2

Mental Chronometry and the Unification of Differential Psychology

Arthur R. Jensen

Mental chronometry is the measurement of cognitive speed. It is the actual time taken to process information of different types and degrees of complexity. The basic measurements are an individual's response time (RT) to a visual or auditory stimulus that calls for a particular response, choice, or decision.

Since at least the time of Sir Francis Galton (1822–1911), the father of differential psychology, it has been hypothesized by him and many others that mental speed is a major aspect of general intelligence. What we now know for sure is that RT can be a highly precise, reliable, and sensitive measure of individual differences. Its relationship to other psychological and ecological variables, however, is a complex affair just recently being explored.

Research on RT has a venerable history. Not only was it the earliest measurement technique used in empirical psychology, but also its scientific use as a measure of individual differences preceded the beginning of experimental psychology by at least half a century. The first published research on RT appeared in astronomy journals. Time, as measured by the Earth's rotation with reference to a star's moment of transit across a hairline in the lens of a telescope, had to be measured as accurately as possible. In 1796 it was accidentally discovered by the Astronomer Royal at the Greenwich Observatory that astronomers showed individual differences in RT to the star's transit across the hairline. So it was decided that each astronomer's RT had to be "corrected" for any given individual's "personal equation," that is, the deviation of the individual's mean RT from the mean RT based on the observations made by a number of astronomers. Before then, it had been assumed that such simple RT was virtually instantaneous. RT was later taken up as a basic tool in experimental psychology. Shortly thereafter, the measurement of individual differences in RT, along with other measures of human capacities, became a subject of interest in its own right to Galton, who tested simple RT on thousands of people (Jensen, 1982b, 1994). Unfortunately, by the early 1900s the purely technical inadequacies of this early work in mental chronometry caused the near demise of this field of investigation, and subsequent developments in mental testing were dominated for nearly a century by the psychometric model based on the famous intelligence test devised by Alfred Binet in 1905. In recent years, however, the premature abandonment of chronometric methods in differential psychology has been rectified by a rapidly burgeoning research literature in this field, particularly related to the nature of intelligence conceived theoretically as the speed and efficiency of information processing (Vernon, 1987).

This chapter explains the important differences between conventional psychometric measurement and mental chronometry and points out the particular advantages of chronometry and its future prospects for the advancement of differential psychology as a natural science.

PSYCHOMETRY AND CHRONOMETRY COMPARED

The practical success of psychometrics is unquestionably one of the triumphs of applied psychology. When nothing more than ordinal measurement is required, there can be little dispute about the practical usefulness of item-based mental tests. These are composed of a number of separate items on which the subject's responses are scored either right or wrong (R/W) or pass/fail (P/F). Given the variation in items' *p* values (the proportion of the normative sample passing the item) and given a range of individual differences in the ability to pass the items, the distribution of total scores (e.g., the number right) constitutes an ordinal scale. An individual's score on such a scale is interpreted in terms of its location in the distribution of scores obtained in some specified group, so the scores are "norm referenced." The interpretation of normative scores is facilitated by various forms of scaling, such as ranking, percentile ranks, standardized scores (e.g., *z*, T, IQ), normalized scores, and various Rasch-type scales.

To suppose that any kind of transformation of the raw scores' rank order represents a true *interval* scale or a *ratio* scale, rather than merely an ordinal scale, depends entirely on an assumption. Plausible and practical though this assumption may be, it remains just an assumption. We have to assume that the distribution of the essential variable, or latent trait, measured by the test has a particular form in the normative population. Psychologists usually assume that the trait has a normal (Gaussian) distribution, so the ordinal score distribution, whatever its form, is mathematically transformed to conform to this assumption. Or items may be specially selected for difficulty level and item intercorrelations that will produce an approximately normal distribution of scores. The transformed or manipulated test scores contain no new information that was not present in the rank-ordered raw scores; the form of the distribution simply reflects the initial assumption.

Ordinal scales have many shortcomings. Without a true *interval* scale, but armed with only our faith in the unproved distribution assumption, we cannot make really meaningful statements about many things we want to know in differential and developmental psychology. For example, there are obvious questions about the form of the population distribution of a given trait, or the form of the growth curve for that trait, and its rate of change across the lifespan. Knowing these things depends on having equal interval measurements throughout the full range of variation in the characteristic of interest. For the same reason, meaningful comparison of the within-group variance between different groups whose score distributions are centered in different ranges of the scale depends on measures having equal units across the whole scale. Otherwise a difference of X points between two scores near the high end of the scale is not assuredly equivalent to a difference of X points near the low end. The precision of covariance and of both the Pearson r and intraclass correlation (but not Spearman's rank correlation) depends on equal-interval measurements of both variates. Without an interval scale the specific form of any functional relationship, as might be shown on a graph, say, in which mental test scores (y axis) are plotted as a function of drug dosages (x axis), provides no dependable information over what could be expressed by the rank correlation coefficient between the x and y variables.

A *ratio* scale, with both a natural zero point and equal intervals, is even less attainable by any plausible assumptions based on item statistics than is an interval scale. Yet a ratio scale is essential for any valid mathematical manipulations of data beyond simple additivity. Without ratio scale properties, multiplicative or ratio properties of the data cannot be known. About 35 years ago, for instance, some psychologists proclaimed that children, on average, acquire one-half of their mental growth potential by four years of age. But psychometrics has no measurement scales that could test this interesting claim. Answering this kind of question about height, or weight, poses no problem at all. It would be scientifically useful if psychologists could determine the functional relationship of various mental measurements to the precisely known growth curves for certain structures of the brain. But our psychometric tests cannot do this meaningfully. At best, they cannot really provide anything more informative than a rank correlation between any mental ability and any metrical property of the brain.

This absence of ratio scales in differential psychology is most unfortunate, as many psychological variables behave multiplicatively, exponentially, or logarithmically in relation to internal and external physical variables, as has been discovered in sensory psychophysics, probably the most advanced branch of psychology where measurement is concerned.

The noted limitations of the scale properties of psychological tests and the claimed advantages of true interval and ratio scales might be dismissed as a trivial issue for most aspects of applied psychometrics, for which reliable ordinality is sufficient for the practical predictive validity of tests. It is not sufficient, however, for the advancement of differential psychology as a natural science, especially the study of individual variation in the domain of cognitive abilities, including the well-established dimensions, such as *g*, verbal, and spatial factors. With only ordinal scales we do not know the true form of the population distribution of each of these different factors or the true amount of variance attributable to each one. Nor can we know or compare their growth curves or their rates of decline with age. The future of reductionist research in this field, which aims to be explanatory, will necessarily be focused on discovering functional relationships between behaviorally measured cognitive abilities and their causal physical properties and processes in the brain. A main scientific purpose of measurement is the discovery and description of how one measured variable is related to some other measured variable. Ideally, and often necessarily, the measurements on *both* sides of the equation should be ratio scales. The physical measurements in brain research per se are of course ratio scales. Arguably the most natural scale for the behavioral measurement of mental activity is *time*, a physical ratio scale of international standardized units.

ADVANTAGES OF MENTAL CHRONOMETRY

Mental chronometry (MC) has two main classes of paradigms: (1) the measurement of an individual's *response time* (RT) to a *reaction stimulus* (RS) that elicits some form of mental activity and (2) the measurement of an individual's *inspection time* (IT), or the minimum length of exposure needed by the subject to discriminate between stimuli that differ on some dimension. MC also includes *derived* measures obtained from mathematical relationships (sums, products, ratios, etc.) between various RTs (or ITs), and these also have the scale properties of physical measurements. Nowadays RT is measured by an electronic apparatus that accurately registers intervals of time in milliseconds (ms). Besides the undisputed virtue that *time* is a ratio scale measurement, what are some of the most general advantages of MC for advancing a true science of differential psychology?

RELIABILITY. RTs are always measured over a number of trials. The internal consistency reliability (e.g., Cronbach's coefficient alpha) of individual differences in the mean RT obtained from a given number of trials can be made as high as may be required for a particular purpose simply by increasing the number of test trials. Reliability coefficients as high as those of most good psychometric tests can be obtained in as few as 20 to 30 trials, taking only a few minutes. The alpha reliability coefficients for different numbers of trials conform near perfectly to the values predicted by the

Spearman–Brown prophecy formula because the essential condition on which the S–B formula depends is perfectly met, i.e., every RS is randomly sampled from the same pool of RSs.

REPEATABILITY. Most chronometric tests can be repeated in identical form over and over again. There is virtually an infinite supply of equivalent forms of a specific test that are truly equivalent across administrations. Practice effects are typically small compared to individual differences; they approach asymptote after a certain number of trials (depending on RS complexity), and they have relatively little effect on the reliability of individual differences across trials or occasions. Repeatability of measurement is a great advantage for a test that is used over an extended period of days, weeks, or months to monitor a behavioral or cognitive effect of a drug or other treatment. Repeatability is also a boon to the study of drug-dosage curves; a given cognitive effect can be functionally related to differing dosages of the drug. Because of this advantage, MC is now of interest to pharmaceutical firms and treatment hospitals, as more and more new drugs unintentionally have side effects on cognitive performance that cannot be monitored repeatedly by ordinary item-based tests.

RANGE OF EQUIVALENCY. Conventional psychometric tests typically have a very narrow range of equivalency compared to chronometric tests. The IQs of low-scoring individuals on a test like the Wechsler Adult Intelligence Scale (WAIS) are based on a largely different set of items than are the IQs of high scoring individuals. Thus because of the limited range of a given item's *p* values for individuals in different segments of the score distribution, strictly speaking the same test cannot be given to low, medium, and high scoring persons. Without Rasch scaling, at least, it is even questionable whether the same variable is being measured in the different ability groups. The same problem applies to children of different ages. Even though a five-year-old and a ten-year-old are given nominally the same test, they have actually been tested on entirely different discriminating items, unless they obtain nearly the same raw score. The range of ability or age within which the same test items are discriminative is remarkably narrow. In contrast, one and the same chronometric test, with a set number of trials, can discriminate as reliably among preschool children as among university students, and among gifted as among mentally retarded children. Moreover, in all of these groups the chronometric measures have shown similar correlations with IQ. The groups differ markedly in mean RT, of course, and one can describe the differences in mathematically meaningful terms. But with ordinary item-based tests given to such diverse groups we could only rank the group means and estimate the statistical significance of their differences. Direct comparisons of ability levels would be meaningless or impossible.

SENSITIVITY OF MEASUREMENT. RT is an extraordinarily sensitive measure, showing reliable individual differences and within-subject differences in the cognitive demands of various elementary tasks that are virtually undetectable by psychometric tests. A classic example is a chronometric analysis that shows how schoolchildren in the first grade perform the simple arithmetic task of adding two single digit numbers (Groen & Parkman, 1972). On each test trial the subject is shown two integers that always sum to values from 0 to 9. The subject responds by pressing one of ten keys labeled with the digits 0 to 9, and the RT is measured in milliseconds. Analysis of the RTs revealed what the children were doing mentally: First they selected the larger (*L*) number in the given pair; then they counted up the smaller (*S*) number (perhaps using their fingers). The RTs measured on the various problems increased as a linear function of *S*, indicating that even simple addition is not merely the unitary recall of a memorized number fact but is a strategic construction. The contrast between this constructive effect and sheer memorization is seen in the finding that when both numbers in the pair are the same, there is no systematic variation in RT. This suggests that the sum of any two identical digits has been memorized as a unit and RT simply reflects the time for retrieval of this item of information from long-term memory. The RT for retrieval averages less than the RT for construction.

It is most interesting that these very same strategic and memorial phenomena are found also in young adult college students, although their RTs average only about one-fourth the RTs of first-grade children. But the college students are still constructing addends from pairs of single digits in the same way as first graders, only much faster. But college-age students are also much faster than young children on every kind of RT. Studies of elementary schoolchildren selected for ability to perform perfectly on simple addition, subtraction, and multiplication problems given as untimed paper-and-pencil tests have shown significant individual differences when RTs are measured on the same problems. There are also consistent mean differences between RTs for addition, subtraction, and multiplication, indicating differences in complexity of processing for the three types of arithmetic (Jensen, 1998a, see references to Jensen & Whang). These pupils' RTs on such simple arithmetic problems predicted their ability in more advanced types of arithmetic problem solving, consistent with the hypothesis that success in complex problem solving depends in part on the speed with which elementary components of the problem can be processed. Indeed a whole psychology of arithmetic cognition could be ferreted out of cleverly designed experiments based on chronometric analysis.

Other evidence of sensitivity is that chronometric measures detect variation in physiological state associated with an individual's metabolic diurnal cycle, changes in body temperature, effects of exercise, stimulant and depressant drugs, medical conditions, and the presence of genes that are risk factors for the development of Alzheimer's disease, such as the apolipoprotein (APOE) e4 allele, even before its cognitive effects are clinically detectable by psychometric tests specifically designed for this purpose (O'Hara, Sommer, & Morgan, 2001).

The sensitivity of RT can also be a disadvantage in that it is a source of variance that acts as a measurement error in studies of individual differences. In studies of *intra*-individual differences, the sensitivity of RT can be taken into account by obtaining repeated measures always at the same time of day and monitoring indicators of physiological state at the time of testing and the time since the last meal, body temperature, drug usage, and time in the menstrual cycle.

THE PSYCHOMETRIC MISCONCEPTION OF MENTAL SPEED

Psychometric measures of mental speed, such as the digit symbol or coding subtest of the Wechsler scales and the clerical checking subtest of the Armed Services Vocational Aptitude Battery, are mentally very easy tests on which virtually all subjects would obtain a perfect score if the tests were not highly speeded. The score is the number of items completed within a given time limit. Such speeded tests have often been included in factor analyses with many other more complex mental tests, such as vocabulary, verbal and figural analogies, problem arithmetic, matrices, and block designs, to name a few. In a hierarchical factor analysis these speeded tests typically show up as rather small first-order factors; they have little variance in common with other tests as shown by the fact that they have smaller loadings than other tests on any of the higher-order factors, least of all on the most general factor, psychometric *g*. This has resulted in a long held and strongly entrenched misconception in psychometrics that mental speed is a minor factor in the abilities hierarchy and has little relevance to higher mental abilities or the *g* factor.

The kinds of tests identifying this psychometric speed factor are decidedly different from the chronometric methods used to measure RT and IT, which behave quite differently from the speeded tests used in psychometrics. RT measured in various chronometric paradigms generally has its largest correlations with the nonspeeded and most highly *g* loaded tests, whereas its lowest correlations are with the most speeded tests like the digit symbol subtest in the Wechsler scales. Moreover, the correlations of various RT measures with each other and with various nonspeeded psychometric tests are generally similar to the correlations among the various subscales of standard test batteries. More generally, we should realize that the traditional distinction between *speed* and *powe*r in describing psychometric tests is strictly a formal distinction. It is a mistake to attribute these purely descriptive terms to categorically different cognitive processes.

STANDARDIZING CHRONOMETRIC METHODS

The study of individual differences in RT originated in astronomy, when extremely precise measurement of individual differences in RT, the so-called personal equation, was critical in measuring the instant a star's transit crossed a hairline in the telescope. The units of time have been standardized throughout the history of MC. Today these units, measured electronically in milliseconds, are the same in all laboratories. What is seldom realized, however, is that the testing conditions for obtaining these measurements in different laboratories are not at all well standardized. This is most unfortunate for the development of a unified science. Under a comparable handicap the physical or biological sciences could not have progressed to their present level. This condition has seemed tolerable where MC is used in experimental psychology, but it will prove a severe hindrance to differential psychology. This is because the former is concerned with the effects of experimentally varying task parameters and measuring the effects on RT *within* subjects, while variation *between* subjects is treated as unwanted error, to be minimized by averaging RTs over a number of subjects or over many test trials in a single subject. Only the direction and relative magnitudes of the experimental effects are of interest. Thus it is not a critical disadvantage that the exact numerical values of RT vary from one lab to another, so long as the relative effects of experimental manipulations are replicable across different labs.

Because differential psychology is concerned with differences *between* subjects, the *absolute* values of RT become important. This calls for standardization of the methods by which RT is measured, unless we limit our uses of chronometry to discovering purely relative effects and performing only correlation analyses, methods for which measures of central tendency and variance are irrelevant. Without standardization MC loses many of its advantages. The failure of one lab to replicate the specific findings of another lab using nominally the same paradigm can be due either to differences between the subject samples or to differences in the test instruments themselves, although both are measuring and comparing, say, simple RT and 2-choice RT to visual stimuli. Unless the same apparatus (or perfect clones), as well as the instructions and the number of practice trials, are used in both labs, a true replication is not possible.

The sensitivity of RT makes for considerable differences when nominally the same variable is measured by different, though equally accurate, apparatuses. The difference arises not in the timing mechanism per se, but in subtleties of the stimulus and response demands of the task. Given the same testing conditions, any significant difference in results should be solely attributable to a difference between the subject samples, not to the conditions of measurement. Regardless of the RT data collected for a particular study, an important element in describing the subject sample (besides the usual descriptors such as age, sex, and education) should consist of descriptive statistics based on, say, at least 20 trials of both simple RT and 2-choice RT measured on the standard RT apparatus. Without such methodological standardization in differential research, the cumulation of archival data from different laboratories is hardly worthwhile. Such fundamental standardization has been essential for progress in the so-called exact sciences, and it is equally important for the advancement of a science of differential mental chronometry. Decisions about the design of standard apparatuses, methods, and procedures that should be required in every chronometric laboratory will need to be worked out and agreed on by an international consortium of researchers in this field. This agreement would also include recommendations for electronically recording and archiving chronometric data from labs using the standardized equipment and procedures. I find it hard to imagine a greater boon to the advancement of differential psychology, with its present aim of discovering how behavioral measurements of cognition are related to the physical properties of the brain.

CHRONOMETRY AS A PRIMARY TOOL FOR RESEARCH ON INTELLIGENCE

The century of progress in the psychometric approach to the study of mental abilities, beginning with Spearman and Binet, has reached a consensus regarding their factor structure. Relatively few factors, or latent variables, account for most of the individual differences variance in practically all psychometric tests. John B. Carroll's (1993) systematic factor analysis of the huge number of test intercorrelations reported in virtually the entire psychometric literature shows that they are best represented by a hierarchical factor structure. Carroll named it the *three-stratum model*. It comprises some forty first-order factors in the first stratum, eight second-order factors in the second stratum, and one factor (psychometric *g*) in the third stratum. The challenge now is to discover the causal basis of the individual differences from which these factors arise. Researchers now want to understand them in terms of cognitive processes and brain physiology. The greatest interest so far is focused on *g*, the most general component of the common factor variance. It is also the most mysterious, as it cannot be characterized in terms of the information content of mental tests or in terms of any observable types of behavior. As its discoverer Charles Spearman noted, *g* is known not by its nature but by the variation in its loadings on a wide variety of mental tests. But psychometric tests with the same *g* loadings are so highly varied in their specific information content and the particular mental skills called for as to defy a unitary classification in lexical terms. The *g* factor itself is best thought of not as a verbally describable mental ability, or even as an ability of any kind, but rather as an aspect of individual differences that causes positive correlations between virtually all measurable cognitive abilities.

The individual assessment of *g* is always problematic, not because *g* is a chimera, but because its psychometric measurement as a factor score is always attached to a *g*-weighted average of a relatively small number of diverse tests. Therefore, the psychometric "vehicles" of *g* also unavoidably carry other factors besides *g*, including variance unique to each test. We can only minimize these sources of non-*g* variance in the obtained *g* factor scores. But because the contamination of *g* factor scores by the vehicles of *g* is unavoidable, this attempt can only be more or less successful for different individuals. Fortunately for research on the nature of *g*, it is unnecessary to have a direct measure of *g* for each individual in a study. One can indirectly determine the correlation of *g* with other psychological and physical variables by the methods of factor analysis or other latent trait models.

The advancement of intelligence research along scientific lines now requires extending its traditional methodology beyond the use of item-based psychometric tests and the factor analysis of the virtually unlimited variety of tests. During the past two decades, chronometric methods have gained prominence in research probing the nature of *g* and other components of psychometric variance. It is now well established that many types of RT and IT are correlated with psychometric *g* and with IQ or other highly *g*-loaded tests. The correlations for single elementary cognitive tasks (ECTs) with RTs in the range from simple RT (about 200 ms) to more complex tasks (not exceeding 2,000 ms in young adults) the correlations with IQ range from about .10 to .50. The general factor extracted from a battery of several diverse ECTs has correlations with the general factor of a battery of psychometric tests (e.g., the Wechsler scales) ranging between .60 and .90. Studies of the RT/IQ relationship based on multiple regression, factor analysis, canonical correlation, and structural equation models suggest that chronometric and psychometric tests have much the same general factor in common. Reviews of the empirical evidence and bibliographic entries to virtually the entire literature on this subject can be found elsewhere (Caryl et al., 1999; Deary, 2000a, b; Jensen, 1982a, b, 1985, 1987a, 1998a, Chap. 8; Lohman, 2000; Neubauer, 1997; Vernon, 1987). So here I will not reiterate the evidence proving that RT and IT are related to *g*. Rather, I shall point out some of the collateral phenomena that have turned up in this field of investigation. Their investigation is important for advancing this line of research. A true theory of *g* and its neural basis will have to account for each of these phenomena, unless future research finally dismisses them as unreliable or as experimental artifacts.

But first let me emphasize that the eventual explanation of *g*, as marvelous an achievement as that might be, is not the main purpose of mental chronometry. Its scope is far wider. It is a general tool for measuring all aspects of cognition. Our conventional psychometric tests, whatever their practical usefulness, are not a higher court to which mental chronometry must appeal for its scientific importance. Chronometric methods have generated a universe of psychological phenomena for study in its own right.

That some of these phenomena happen to be related to psychometric test scores is simply a fortunate discovery, helping us understand individual differences in more functionally analytic terms than is possible with the factor analysis or multidimensional scaling of item-based tests. We recognize, of course, that these psychometric methods have served a necessary taxonomic purpose in describing the whole domain of psychometric abilities in terms of a quite limited number of latent variables.

FUNDAMENTAL FINDINGS IN THE RELATIONSHIP OF CHRONOMETRICS TO PSYCHOMETRIC *g*

Speeded Psychometric Tests

The RT–*g* correlation is not in the least explained by the time limits or speed instructions given to the subjects taking the mental tests. In fact, the types of tests that are usually the most speeded, such as clerical checking and digit-symbol coding tests, have lower correlations with RT than do so-called power tests, in which subjects are encouraged to attempt all the items and to take all the time they need.

Tests' g *Loadings*

Tests with larger *g* loadings generally show higher correlations with RT, indicating that *g* is the main psychometric factor in the RT–IQ correlation.

Complexity of the RT Task

The absolute size of the RT–IQ correlation (which is always a negative *r*) generally has an inverted-U-shaped relationship to the complexity of the RT task. Simple RT (i.e., one stimulus–one response) with RTs of about 300 ms for young adults shows *small* correlations (−.10 to −.20); moderate tasks (RTs around 500–900 ms) show *moderate* correlations of (−.40 to −.50); and difficult RT tasks (above 1200 ms) show *small* correlations (−.20 to −.30). One hypothesis proposed to explain this phenomenon holds that the simplest RT tasks have a smaller cognitive component relative to a larger perceptual-motor component, which does not reflect *g*. As the RT task demands are increased in cognitive complexity beyond some optimal point, a wider range of individual differences in an increasingly greater variety of performance strategies comes into play. These include task-specific factors that are uncorrelated with psychometric factors and therefore attenuate the RT–*g* correlation. Also, when task complexity increases to the point that response errors become a reliable source of individual differences, fewer subjects are processing the RT task in the same way. Interestingly, those forms of both RT and IT tasks that are the most liable to allow subjects to adopt different strategies show the weakest correlations with IQ. Evidently it is the sheer speed of processing, rather than the subject's choice of a strategy, that is most related to *g*.

Because we are often without an independent interval scale of task complexity, task complexity is often measured by RT itself. Such RT measures on simple tasks, though differing only in tens of milliseconds (i.e., time intervals below the threshold of visual or auditory detection), have considerable subjective validity as measures of task complexity. This was shown when a group of university students was asked simply to rank the complexity (or difficulty) of fourteen different items in a Semantic Verification Test (SVT, described in the following section). Their subjective ranking of item complexity, from least complex (=1) to most complex (=14), correlated +.61 with the item's average RTs obtained in another university sample (Paul, 1984; Jensen, Larson, & Paul, 1988). It could well be that RT provides the best measure of item complexity and could be used in the process of item selection in the design of ordinary paper-and-pencil tests for children. Simple test items can be scaled on a ratio scale of difficulty according to their average RTs obtained in a group of bright university students who can answer the items without error. Reliable discrepancies between the item *p* values for children and the item RTs for university students would indicate that *p* and RT are not scaling item difficulty (or complexity) on the same dimension. I predict, however, that this would very seldom occur.

Correlation Trade-Off and Convertibility Between RT and Error Responses

As RT tasks increase in complexity, there is a rise in response errors. The correlation between RT and IQ *decreases* with a rise in response errors, whereas the correlation between response errors and IQ *increases*. This reciprocal trade-off suggests a breakdown in information processing at higher levels of task complexity. The point of breakdown on the continuum of difficulty or complexity and the resulting response error determine the correlation of single test items (scored pass/fail) with IQ.

Untimed psychometric tests based on right/wrong scoring of items with little or no prior-learned knowledge, such as the Raven matrices and number series tests, are an example of this; the average item scores (*p* values) reflect differences in item complexity or difficulty. If items are so easy that nobody misses them (i.e., all item *p* values = 100%), differences in their difficulty levels can still be determined by measuring the RTs for solving the items.

The convertibility between item RTs and item error rates can be shown by means of a simple Semantic Verification Test (SVT) (Paul, 1984). Each item in the test consists of a simple statement about the relative positions of just the three letters **A, B, C**. There are 14 different statements, such as **B after A**, or **B not before A**, or **B between A and C**, etc., with a total of 84 presentations. Immediately following a 3-second presentation of one of these statements on the display screen, three letters (e.g., **A C B)** are presented simultaneously in an order that either affirms or disaffirms the statement.

The subject, instructed to respond as accurately and quickly as possible, presses one of two pushbuttons labeled **YES** or **NO**. The SVT was very easy for Berkeley undergraduates whose average rate of response errors over 84 test trials was 7%. But their mean RTs on the 14 SVT items varied widely, between 600 and 1,300 ms. Obviously the items differ in complexity or difficulty. (The correlation of subjects' mean RTs with scores on the Raven Advanced Progressive Matrices was −.45 in Berkeley undergraduates.)

To obtain reliable measures of variation in item difficulty among the fourteen SVT conditions measured as the *p* values of the SVT items, these items had to be given to schoolchildren (ages 8 and 9 years) as an *untimed* paper-and-pencil test, with an average item *p* value of 82%. The children's *p* values on the fourteen SVT items had a rank-order correlation of −.79 with the mean RTs of the corresponding SVT items in the adult sample. The more difficult an SVT item was for the children, the greater was its average RT for university students. Thus an index of item difficulty (*p*) for average third-grade schoolchildren is convertible into processing time (RTs) for university students all in the top quartile of the nationally normed IQ.

Primary versus Derived Measures in Chronometric Paradigms

Primary measures are the central tendency (*mean* or *median*) of an individual's RTs over a given number (*n*) of trials. Derived measures are (1) the *standard deviation* of an individual's RTs over *n* trials (RTSD), (2) the *intercept* of the regression of mean RTs on task difficulty, and (3) the *slope* of the linear regression relating the individual's mean RT on two or more tasks to their differences in complexity (hence in RT). The slope parameter is a key feature of three classic RT paradigms: the Hick paradigm (linear slope of RT over four levels of complexity measured in bits), the Saul Sternberg paradigm (linear slope of RTs over 1 to 5 or more digits to be scanned in short-term memory), and the Posner paradigm, where the slope is the difference between only two means (Name Identity RT *minus* Physical Identity RT). These slope parameters are of considerable theoretical interest, as the steepness of the slope is a prima facie measure of the *rate* of information processing as a function of increasing information load. An index of *skewness* of an individual's RT distribution over *n* trials is another derived measure that has more recently become of interest in connection with the "worst performance rule" (discussed later).

The derived measures typically show lower correlations with IQ than do the primary measures, which at least in the case of the slope parameter is definitely contrary to the theoretical prediction. But the prima facie evidence against the theoretical prediction that the slope parameter should be correlated (negatively) with IQ at least as much if not more than the mean RT was a premature and technically mistaken judgment. Two statistical artifacts work against the overly simple analysis typically used to test the prediction, namely, a simple (zero-order) correlation between slope and

IQ: (1) the low reliability of the slope measurement and (2) the intercept measurement acts as a suppressor variable in the slope–IQ correlation (because the intercept and slope share the same measurement errors but in opposite directions). These unwanted statistical effects are not intrinsic to the theoretical prediction, but they can be taken into account by an appropriate statistical analysis based on disattenuating the slope measure and partialling out the intercept from the IQ–slope correlation. When such an analysis is applied to the Hick paradigm, the theoretical prediction of the slope–IQ correlation is significantly borne out (Jensen, 1998b).

It should always be remembered that any derived measures, if based on difference scores, $X - Y$, will have lower reliabilities than either X or Y to the degree that X and Y are correlated with each other. This is sometimes forgotten in studies of individual differences in the Posner paradigm and other difference scores such as the difference between choice RT and simple RT. Not taking proper account of reliability in different derived measures is often the reason why derived scores in RT studies result in weaker correlations with external variables like IQ than do the primary RT variables.

The Problematic Meaning of Inter-Trial Variability of RT

Inter-trial variability, also referred to as intra-individual variability, is measured as the standard deviation of an individual's RTs over n trials, abbreviated RTSD. Its interest inheres in the hypothesis that RTSD measures individual differences in "neural noise" or the result of random effects in the transmission of information in the brain, and that the amount of neural noise is a causal factor in intelligence differences. RTSD is negatively correlated with IQ in various paradigms to at least the same degree as the median RT, even though RTSD usually has somewhat lower reliability than RT, so that when all of the statistical parameters of the RT performance are corrected for attenuation, RTSD has the largest correlation with IQ. It therefore commands attention in the chronometric study of cognitive differences.

RTSD has two problematic aspects, as yet unresolved. First is the question of redundancy of the mean RT and RTSD. The near-perfect constancy of the proportionality between the mean RT and RTSD, measured as the coefficient of variation ($C_V = \sigma/\mu$), both for individuals and for different tasks is well established. It implies a perfect correlation between RT and RTSD, corrected for measurement error. Therefore it is mysterious that these two measures do not have the same correlation with IQ and that they show significant interactions with race and sex differences (Jensen, 1992a). Furthermore, analysis of several sets of median RT and RTSD showed that the true-score correlation between the two variables is very high (averaging +0.81), but that still leaves a significant 36% of the variance that the two measures do not have in common. This noncommon variance could result from the fact that all these analyses were based on median RT over n,

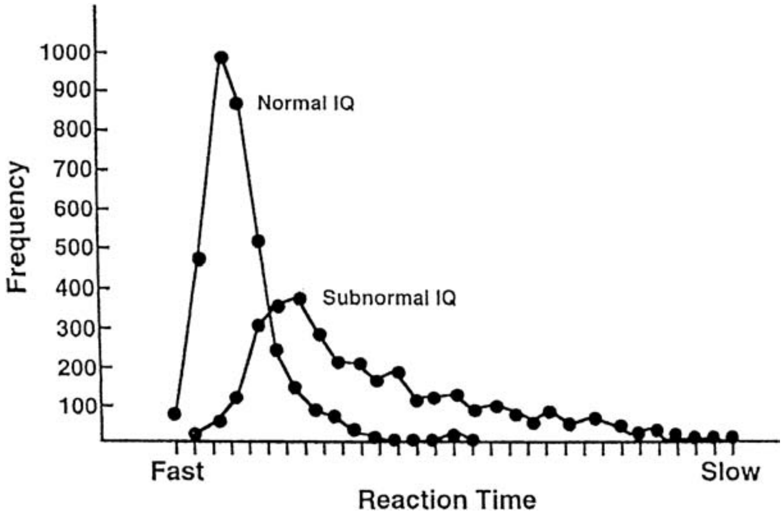

FIGURE 1. Distributions of reaction times of individuals with normal and subnormal IQs. (From Baumeister, 1998, p. 260, with permission of Ablex.)

not the mean RT. Because the RT distribution is always positively skewed, the mean is always somewhat larger than the median. But it has not yet been determined whether a perfect true-score correlation exists between the *mean* RT and RTSD. If there is a perfect correlation, a purely statistical theory could account for it, as follows: (1) Every individual, at a given time, has a physiological limit for the speed of reaction, determined by the minimum times for sensory transduction of the stimulus and the nerve conduction velocity and synaptic delays going to and from the relevant sensory and motor regions of the brain. (2) On a given RT task, the range of individual differences in the physiological limit is much smaller than the range of individual differences in the central tendency (particularly the mean) of RTs measured over many trials. (3) The location of the mean RT, therefore, is determined by the distribution of RT deviations above the physiological limit. (4) Because these deviations can only go in one direction, their distribution is skewed to the right. (5) Whatever causes the variable deviations in RTs thus has three perfectly correlated effects on the first three moments (mean, *SD*, and skew) of the individual's RT distribution. Empirically, over many trials, the correlations among individual differences in the mean RT, the RTSD, and skewness would approach unity. Theoretically, then, the parameters of an individual's RT distribution would all result from the individual's physiological limit plus positive deviations of RT from that limit. This deviation phenomenon would be more or less equally reflected by any one of these moments of the individual's RT distribution. This phenomenon is illustrated in Figure 1.

This hypothesis reduces the problem of explaining the RT–IQ relationship to that of explaining the cause(s) of the RT deviations above threshold. Is it "neural noise," implying true randomness, in which individuals would differ? Or could it be a regular oscillation in neural receptivity, the periodicity of which differs across individuals? A regular oscillation of excitatory potential would simply appear to be random if on each test trial the experimenter-controlled presentation of the reaction stimulus (RS) was seldom synchronized with the individual's period of oscillations above and below the threshold of excitation for the given stimulus. We know that increasing the intensity of the RS correspondingly decreases both the mean RT and the RTSD, indicating that the threshold for the activation of a response operates as a gradient or wave, not as dichotomous on/off levels of stimulus receptivity.

ANOTHER MEASURE OF RT VARIABILITY. For researching this hypothesis, RTSD is not an ideal measure of individual variation in RT across trials. It is liable to include any systematic variation or trend in RTs across trials, such as a practice effect. It would be more desirable to measure an individual's RT deviations across trials in a way that would determine if successive deviations look as if they were produced by a random numbers generator, given the lower limit and the mean of the individual's RT distribution.

Such a measure of random variability, that does not reflect systematic trends in the trial-to-trial RT measures, is provided by Von Neumann's (1941) *mean square successive difference* (MSSD), or its square root. The MSSD is defined as $\delta^2 = [\Sigma(X_i - X_{i+1})^2/(n-1)$, where X_i and X_{i+1} are all sequentially adjacent values (e.g., RTs on Trials 1 and 2, 2 and 3, etc.) and n is the number of trials. It is most commonly used in time series analysis in economics, where it is desirable to distinguish between random fluctuations and systematic trends in financial data. The *Von Neumann ratio* ($R = \delta^2/\sigma^2$) provides one of the strongest statistical tests of randomness in a series of n numbers. [The chance probabilities (p) of R for different values of n are given by Hart (1942).] Although this statistic can indicate randomness of RTs, it cannot, of course, distinguish between randomness due to neural noise and randomness due to asynchrony between a regular oscillation in neural excitatory potential and the intervals between presentations of the RS. That distinction would have to be discovered experimentally by pacing test trials to determine if the subject's minimal RTs can be systematically synchronized in accord with a regularly fluctuating oscillation of neural excitatory potential.

The "Worst Performance Rule"

This RT phenomenon was named by Larson and Alderton (1990), who defined it as follows: "The worst RT trials reveal more about intelligence than do other portions of the RT distribution." Their quite robust finding, based on Navy recruits, was replicated with college students on different

RT tasks (Kranzler, 1992); the phenomenon is also observed in comparing persons with relatively low and high IQs (Jensen, 1982a). However, a study by Salthouse (1998) based on very heterogeneous age groups (18 to 88 years) did not show the worst performance rule (to be discussed later).

The analysis for demonstrating the phenomenon consists of rank ordering each individual's RTs on every trial from fastest to slowest RTs and, *within* each rank, obtaining the correlation between the individual's RTs and ability measures (e.g., IQ). The RT–IQ correlations are seen to increase monotonically from the fastest to the slowest RT trials.

This finding, however, appears not to be a new, independent RT phenomenon. It is best viewed as a statistical consequence of the RT variance phenomena described in the preceding section. Individual differences are least in the smallest RT deviations above a physiological limit, and there is an increasing variance of individual differences for larger deviations. The phenomenon is most clearly seen in comparing groups of normal and mildly retarded young adults on simple RT, shown in Figure 2. Even within a normal group of young adults (Navy recruits) there is a monotonically increasing coefficient of variation ($C_V = SD/\text{mean}$), going from the fastest to the slowest RTs (e.g., Larson & Alderton, 1990, Table 4). (The same phenomenon is clearly seen in the study by Salthouse, 1998, Table 1.) Consequently, the larger deviations have less restriction of range, therefore higher reliability and higher correlation with individual differences in IQ. The coefficients of variation across the RT ranks going from the fastest to the slowest RTs, in fact, were correlated .998 with the RT–IQ correlations within the ranks. Therefore the essential phenomenon calling for theoretical explanation is not the derivative worst performance rule itself, but the fact that higher IQ subjects have consistently smaller RT deviations above their physiological limit than do lower IQ subjects. The more basic question is not yet answered: What causes individual differences in the magnitude of these intra-individual RT deviations? The relationship of the various RT parameters (mean, median, *SD*, *MSSD*, skew) to IQ and psychometric *g* all derive from this one fundamental phenomenon.

Although the RT data per se in the study by Salthouse (1998) show essentially the same features as those in other studies, the Salthouse results differ markedly from the others by not conforming to the worst performance rule with respect to ability. Going from the fastest to the slowest RT, the correlations between RT and scores on various cognitive tests (with age partialled out) show no upward trend. And there is a marked downward trend in the correlations between age and RT, going from fast to slow RT. Salthouse (1998, p. 165) attributes this discrepancy between his and the other studies to several method differences – in the RT tasks, the range of RTs elicited, the types of psychometric tests, the subjects' ages, the number of practice trials, and other procedural differences. So many variations simply rule out any possibility of a specific explanation for the

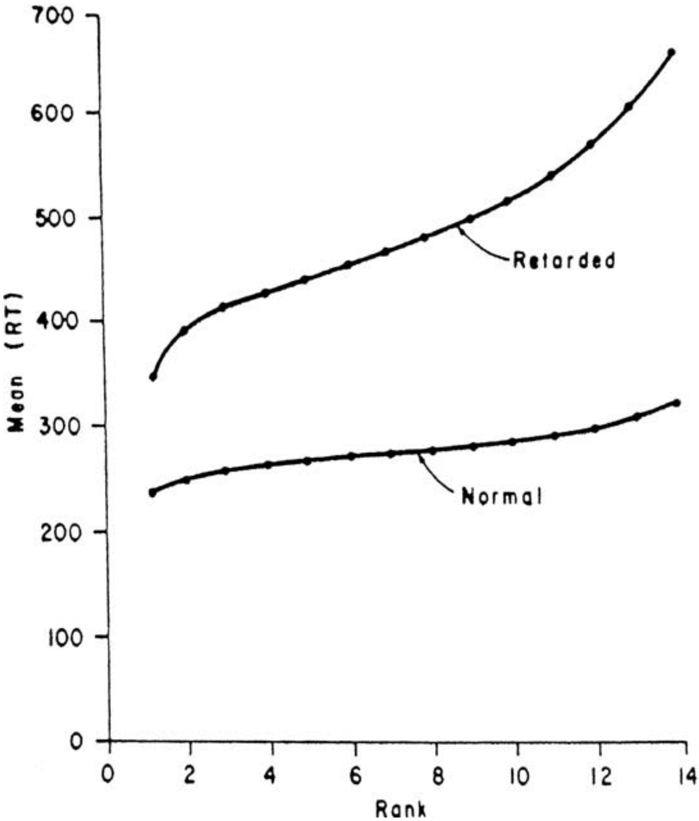

FIGURE 2. Mean simple RT plotted as a function of rank order (from fastest to slowest) of each individual's RTs, for groups of young adults with normal intelligence (mean IQ 120) and with mental retardation (mean IQ 70). (From Jensen, 1982a, p. 291, with permission of Springer-Verlag.)

discrepant results. Each of these studies appears methodologically sound and the results in every instance must be taken seriously, yet each study is so unique methodologically that they can scarcely be regarded as attempts to replicate the same phenomenon. So the worst performance rule is not brought into question, but the limits of its generality is questioned. The importance of true replications of research findings emphasizes the need for standardizing RT apparatuses and procedures in all laboratories engaged in chronometric research.

Working Memory (WM) and Speed of Processing (SP)

Memory is a crucial phenomenon in normal cognition. However, it is not a unitary construct. Stimuli (i.e., information) must be preserved in the neural processing system after their physical presence has ceased, and they must be held long enough in short-term memory (STM) for other processing to occur. If the information input is at all complex and is needed for getting on with the task, it needs to be processed into long-term memory (LTM). That is one of the functions of *working memory* (WM), which is involved in many reasoning tasks and has been called the "mind's scratch pad." WM is a hypothetical ability that (1) rehearses information in STM for storage in LTM, or (2) encodes or transforms information, or (3) simultaneously does 1 or 2 (or both) while processing newly arrived information from the sensorium or retrieved from LTM (Baddeley, 1986). Backward memory span, for example, engages WM capacity more than does forward digit span; the same is true for arithmetic problem solving as compared with mechanical arithmetic. The elements of a problem must be held in WM long enough, or retrieved from the LTM store of past acquired information and cognitive skills, to achieve solution. The *capacity* of WM refers to the quantity of information it can juggle simultaneously without becoming overloaded, causing a breakdown in processing due to the rapid decay of STM traces and the consequent loss of information.

Quite simple laboratory measures of WM have remarkably high correlations with IQ, and it has even been claimed that psychometric g (or fluid intelligence, g_f, which is highly correlated with g) is little, if anything, other than WM capacity. It is hard, however, to evaluate this seeming identity between WM and g. It may be a matter of giving different names to the same construct, as many of the tests of WM are indistinguishable from the highly g-loaded items in psychometric tests. There is no sound basis for pitting WM against mental processing speed as the more fundamental explanation of g. Both constructs – WM and processing speed – are theoretically necessary. The essential question concerns how the two constructs are related. It is a fact that RT derived from simple paradigms is at least as correlated with tests of WM as with nonspeeded g-loaded psychometric tests. RT derives its correlation with various psychometric tests almost entirely through their mutual g loading; when g is statistically removed from a test battery, it has a near-zero correlation with RT. The same is true for WM.

Kyllonen (1993) tested 202 college students on nine diverse WM measures composed either of verbal, numerical, or spatial content and scored as the percentage of correct responses; he also measured 2-choice reaction time (CRT): subjects were presented an alphanumeric stimulus that was either preceded or followed by an asterisk (e.g., *7) and they indicated as quickly as possible which side the asterisk was on by pressing one of two keys positioned 5 inches apart on the left- and right-hand sides of the response console. The average correlation (reflected) between CRT and each of the nine WM tests is .32; the average of all the correlations among just the WM tests is .45. This small difference (.45 − .32) would likely vanish if a slightly more complex CRT paradigm were used. The RT–IQ correlation is increased by including some demand on WM in the RT task. This is done with a *dual task* paradigm, which interposes a different RT task between the first reaction stimulus (RS_1) and the response to it (RT_1), thus: $RS_1 \rightarrow RS_2 \rightarrow RT_2 \rightarrow$ cue for $RS_1 \rightarrow RT_1$, where $RS_2 \rightarrow RT_2$ is the interposed task. Both RT_1 and RT_2 are lengthened by this demand on WM, and both RT_1 and RT_2 show larger correlations with *g* than when either task is presented alone (Jensen, 1987b, pp. 115–118). Thus both processing speed and WM are essential components of individual differences in *g*.

A plausible working hypothesis of the RT–WM correlation is that information processing speed *amplifies* the capacity of WM by a multiplicative factor in which there are consistent individual differences. Here is a brief summary of the points I have elaborated on elsewhere (Jensen, 1982b, 1992b, 1993): (1) The conscious brain typically acts as a *single-channel* processor with *limited capacity*, (2) this restricts the amount of information that can be dealt with simultaneously and the *number of operations* that can be simultaneously performed on it, (3) there is a *rapid decay of information* in STM, which limits the time allowed for manipulating the input or consolidating new information into LTM by rehearsal, (4) overloading the capacity of WM results in a *breakdown in processing*, i.e., some loss of information essential for correctly responding to the task, (5) a faster *speed of processing* allows more operations to be performed on the input per unit of time, thereby increasing the chances of reaching a successful response before the point of overload and breakdown due to loss of information, (6) because of individual differences in speed of processing, a series of novel tasks of increasing *complexity* will show corresponding individual differences in the point of breakdown on the complexity continuum, (7) psychometric tests with items scored right/wrong depend on the complexity continuum (item *p* values) for measuring *g*, (8) therefore, individual differences in speed of processing and its amplification of WM capacity are the cause of psychometric *g*. The specific neural mechanisms involved are not yet known.

Brinley Plots and the Generality of Processing Speed

Differential psychology is mainly concerned with individual differences. But aggregated data, such as mean differences between groups selected to differ on a given trait, afford an essential tool for discovering the common features of the group difference, which consists simply of aggregated individual differences. By aggregating the measurements of many individuals one can distinguish the particular variable of interest from the "noise" caused by other, usually unknown and probably unique, sources of individual variation.

The aggregation principle has been most informative in recent chronometric research studies using a graphical method known as a *Brinley plot*. Originally used in the study of cognitive aging (Brinley, 1965), it consists

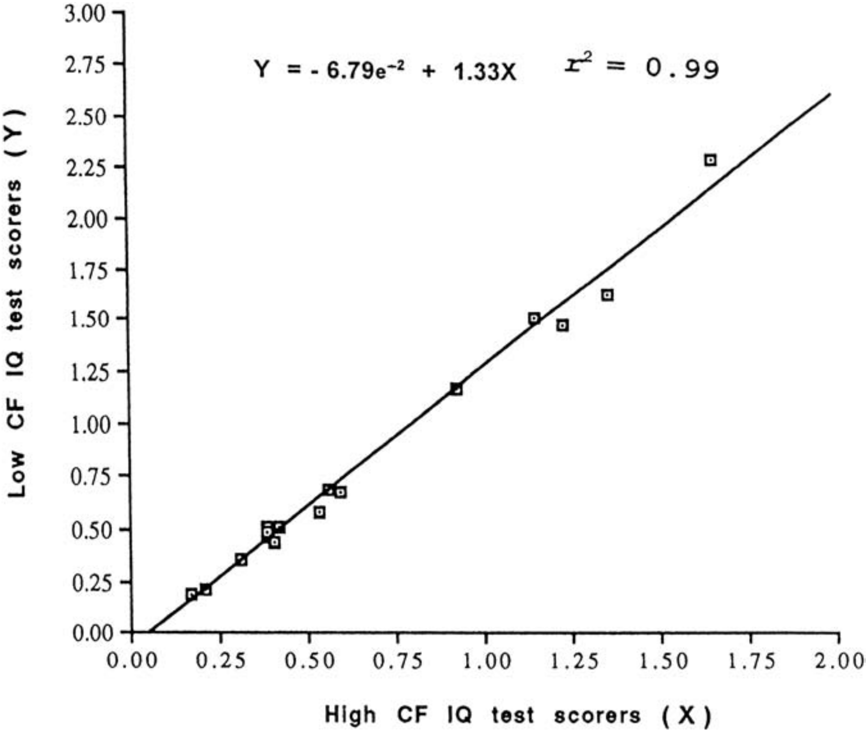

FIGURE 3. Brinley plot of processing speed measures (in seconds) on 15 different RT tasks given to adults in the lower (Low CF IQ) and upper (High CF IQ) halves of the distribution of scores on the Cattell Culture Fair Intelligence Test. The data points are well fitted by the linear regression ($r^2 = 0.99$). (From Rabbitt, 1996, with permission of Ablex.)

of a bivariate plot of the RT means for each of a number of diverse RT tasks in two selected groups (e.g., low IQ and high IQ). One group is plotted on the x axis, the other on the y axis, and the regression line of y on x goes through the bivariate data points. If the contrasted groups should differ in processing strategies on the various tasks, indicating an interaction between groups and tasks, the plotted bivariate means fall off the regression line. The goodness of fit of the RT means to the regression line is indicated by r^2_{xy}, i.e., the proportion of variance in one variate predicted by the other.

An example of a Brinley plot is given by Rabbitt (1996). Cattell's Culture Fair Test of IQ was given to adults who then were divided into the lower and upper halves (called Low CF IQ and High CF IQ) of the total distribution of CF test scores. They also took fifteen chronometric tasks with quite diverse but simple cognitive demands. Figure 3 shows a Brinley plot of the mean RTs on the fifteen tasks. All the data points closely fit a linear function. The squared correlation ($r^2 = .99$) between the RTs of the High and Low IQ groups indicates that 99% of the variance in the fifteen data points of the Low IQ group is predicted by the data points of the High IQ group (and vice versa). The slope of the regression line is indicated by the raw regression coefficient of 1.33, which approximates the average ratio of Low IQ RTs/High IQ RTs across all of the 15 tasks. (The standardized regression coefficient is $r = \sqrt{.99}$.) Rabbitt (1996) interpreted this result as evidence that individual differences in CF test scores (which are highly *g*-loaded) "facilitate all decisions [in the various RT tasks] in close proportion to the times needed to make them, irrespective of their durations (relative difficulty) and of the qualitative nature of the comparisons, and so of the mental processes, that they involve" (p. 79). RT increases *multiplicatively* with task complexity in direct proportion to the number of operations or processing steps involved in the task.

Although a Brinley plot reflects the large global factor (probably *g*) that both the psychometric and chronometric variables have in common, Rabbitt notes that the plot does not capture the fine grain of variation between specific RT tasks. Any given task may differ in the simple ratio of the means of the contrasted groups, thus departing from the common regression line (i.e., the average ratio for all of the RT tasks). Granted this relative insensitivity of Brinley plots for highlighting reliable task specificity (i.e., its interaction with group differences on a second variable such as IQ), it is the multiplicative or *ratio* property, not the additivity, of task differences that is the seminal discovery. It would have been impossible to discover, much less prove, this ratio property of task difficulty without chronometric methods, as they have the theoretical benefit of a true ratio scale. With psychometric test scores, on the other hand, ratios and proportions are meaningless.

Other examples of the Brinley-plot phenomenon are also displayed in Rabbitt's 1996 article and in other chronometric studies of group differences, particularly changes in cognitive abilities across the lifespan. Brinley plots all look much alike, indicating the broad generality of processing speed across a wide variety of elementary cognitive tasks (ECTs) for various kinds of group differences. In every study, the RTs of the slower group are predicted by a single constant multiplier of the corresponding RTs of the faster group. The correlation (predictive validity) is typically in the high .90s. Studies of mental development have compared RTs of children in different grades in school (Fry & Hale, 1996; Hale 1990; Hale & Jansen, 1994; Kail, 1991a, b). Academically gifted 13-year-old students were compared with age-matched average children and with university students on eight RT tasks (Cohn, Carlson, & Jensen, 1985), resulting in Brinley plots averaging a correlation of .96. Studies of cognitive aging used Brinley plots to compare adult groups of different ages (Cerella, 1985; Cerella & Hale, 1994). Brinley plots of RT differences showing the typical global effect of differences in processing speed have also been found in contrasting the following conditions with control groups: brain injury, multiple sclerosis, and clinical depression (references in Myerson et al., 2003). Changes or differences in ability associated with cognitive development, cognitive aging, health conditions, giftedness, and IQ differences at a given age all reflect global differences in speed of processing in a wide variety of RT tasks.

The impressively thought-out article by Myerson et al. (2003) provides the most sophisticated theoretical and quantitative development of this global speed of processing phenomenon. It will prove heuristic to hypothesize that this same global process is the basis of *g* and affects every form of information processing encountered by individuals throughout life.

What ultimately needs to be discovered is the physical basis of differences in cognitive processing speed. Current research based on positron emission tomography (PET scan) and functional magnetic resonance imaging (fMRI) have proven valuable in discovering the specific regions of brain localization for certain cognitive functions, including the areas of cortical activation (mainly in the frontal lobes) associated with performance on high *g*-loaded tests (Duncan et al., 2000; Thomson et al., 2001). Of course, it is important to determine whether the very same cortical areas are activated in performance on the general factor of various chronometric tasks. But the next step in achieving a complete physical account of the causal mechanisms involved in *g* must go beyond studies of brain localization. It must eventually deal with the neural networks in the activated areas on the brain indicated by PET and fMRI. Research strategy in this frontier, similar to the research strategy in particle physics, calls for experimentally testing hypotheses about the known neurophysiological processes that could account for specific behavioral manifestations of *g*, as measured under standardized laboratory conditions. For the reasons outlined earlier in this chapter, I believe that the methods of mental chronometry should prove to be a most valuable research tool for advancing toward this ultimate goal.

References

Baddeley, A. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baumeister, A. A. (1998). Intelligence and the "personal equation." *Intelligence, 26*, 255–265.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy in the elderly. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging, and the nervous system* (pp. 114–149). New York: Springer-Verlag.

Carroll, J. B. (1993). *Human cognitive abilities*: *A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Caryl, P. G., Deary, I. J., Jensen, A. R., Neubauer, A. C., & Vickers, D. (1999). Information processing approaches to intelligence: Progress and prospects. In I. Mervielde, I. Deary, F. de Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 181–219). Tilburg, Netherlands: Tilburg University Press.

Cerella, J. (1985). Information processing in the elderly. *Psychological Bulletin, 98*, 67–83.
Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life span. *Acta Psychologica, 86*, 109–197.
Cohn, S. J., Carlson, J. S., & Jensen, A. R. (1985). Speed of information processing in academically gifted youths. *Personality and Individual Differences, 6*, 621–629.
Deary, I. J. (2000a). *Looking down on human intelligence: From psychometrics to the brain*. Oxford, UK: Oxford University Press.
Deary, I. J. (2000b). Simple information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 267–284). Cambridge, UK: Cambridge University Press.
Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzong, H., Ahmed, A., Newell, F. N., & Emslie, H. (2000). A neural basis for intelligence. *Science, 289*, 457–460.
Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241.
Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.
Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*, 653–663.
Hale, S., & Jansen, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science, 5*, 384–389.
Hart, B. R. (1942). Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics, 13*, 207–214.
Jensen, A. R. (1982a). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 93–132). Hillsdale, NJ: Erlbaum.
Jensen, A. R. (1982b). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.
Jensen, A. R. (1985). Methodological and statistical techniques for the chronometric study of mental abilities. In C. R. Reynolds & V. L. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 51–116). New York: Plenum.
Jensen, A. R. (1987a). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 101–175). Norwood, NJ: Ablex.
Jensen, A. R. (1987b). The *g* beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87–142). Hillsdale, NJ: Erlbaum.
Jensen, A. R. (1992a). The importance of intraindividual variability in reaction time. *Personality and Individual Differences, 13*, 869–882.
Jensen, A. R. (1992b). The relation between information processing time and right/wrong responses. *American Journal on Mental Retardation, 97*, 290–292.
Jensen, A. R. (1993). Why is reaction time correlated with psychometric *g*? *Current Directions in Psychological Science, 2*, 9–10.
Jensen, A. R. (1994). Reaction time. In R. J. Corsini (Ed.), *Encyclopedia of Psychology*, 2nd ed. (Vol. 3, pp. 282–285). New York: Wiley.
Jensen, A. R. (1998a). *The* g *factor*. Westport, CT: Praeger.

Jensen, A. R. (1998b). The suppressed relationship between IQ and the reaction time slope parameter of the Hick function. *Intelligence, 26*, 43–52.

Jensen, A. R., Larson, G. E., & Paul, S. M. (1988). Psychometric *g* and mental processing speed on a semantic verification test. *Personality and Individual Differences, 9*, 243–255.

Kail, R. (1991a). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490–501.

Kail, R. (1991b). Development of processing speed in childhood and adolescence. *Advances in Child Development and Behavior, 23*, 151–185.

Kranzler, J. H. (1992). A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences, 13*, 255–261.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *Journal of General Psychology, 120*, 375–405.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*, 309–325.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 285–340). Cambridge, UK: Cambridge University Press.

Myerson, J., Zheng, Y., Hale, S, Jenkins, L., & Widaman, K. (2003). Difference engines: Mathematical models of diversity in speeded cognitive performance. *Psychonomic Bulletin and Review*, *10*, 262–288.

Neubauer, A. C. (1997). The mental speed approach to the assessment of intelligence. In J. S. Carlson, J. Kingma, & W. Tomic (Eds.), *Advances in cognition and educational practice: Reflections on the concept of intelligence* (Vol. 4, pp. 149–173). Greenwich, CT: JAI Press Inc.

O'Hara, R., Sommer, B., & Morgan, K. (2001). Reaction time but not performance on cognitive tasks identifies individuals at risk for Alzheimer's disease: A preliminary report. Unpublished manuscript, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford University, Stanford, CA.

Paul, S. M. (1984). Speed of information processing: The Semantic Verification Test and general mental ability. Unpublished Ph.D. dissertation, University of California, Berkeley.

Rabbitt, P. (1996). Do individual differences in speed reflect "global" or "local" differences in mental abilities? *Intelligence, 22*, 69–88.

Salthouse, T. A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26*, 153–166.

Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V.-P., Huttunen, M., Longqvist, J., Standertkjold-Nordenstam, C.-G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C. I., & Toga, A. W. (2001). Genetic influences on brain structure. *Nature: Neuroscience, 4*(No. 12), 1253–1258.

Vernon, P. A. (Ed.) (1987). *Speed of information processing and intelligence* (pp. 101–175). Norwood, NJ: Ablex.

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics, 14*, 378–388.