1-1-1987

4. The g Beyond Factor Analysis

Arthur R. Jensen
University of California - Berkeley

4 The g Beyond Factor Analysis

Arthur R. Jensen
University of California, Berkeley

The problem of *g*, essentially, concerns two very fundamental questions: (1) Why are scores on various mental ability tests positively correlated? and (2) Why do people differ in performance on such tests?

SOME DEFINITIONS

To insure that we are talking the same language, we must review a few definitions. Clarity, explicitness, and avoidance of excess meaning or connotative overtones are virtues of a definition. Aside from these properties, a definition per se affords nothing to argue about. It has nothing to do with truth or reality; it is a formality needed for communication.

A *mental ability test* consists of a number of *items*. An *item* is a task on which a person's performance can be *objectively* scored, that is, classified (e.g., "right" or "wrong," 1 or 0), or graded on a scale (e.g., "poor," "fair," "good," "excellent," or 0, 1, 2, 3), or counted (e.g., number of digits recalled, number of puzzle pieces fitted together within a time limit), or measured on a ratio scale (e.g., reaction time to a stimulus or the time interval between the presentation of a task and its completion). *Objectively* scored means that there is a high degree of agreement between observers or scorers or pointer readings in assigning a score to a person's performance on an item.

An item measures an *ability* if performance on the item can be objectively scored such that a higher score represents better performance in the sense of being more accurate, more correct, quicker, more efficient, or in closer conformance to some standard—regardless of any value judgment concerning the aesthetic, moral, social, or practical worth of the optimum performance on the particular task. An item measures a *mental* (or cognitive) ability if very little or none of the individual differences variance in task performance is associated with individual differences in physical capacity, such as sensory acuity or muscular strength, and if differences in item difficulty (percent passing) are uncorrelated with differences in physical capacities per se.

In order for items to show individual differences in a given group of people, the items must vary in difficulty; that is, items without variance (0% or 100% passing) are obviously nonfunctional in a test intended to show individual differences. A test, like any scientific measurement, requires a standard procedure. This includes the condition that the *requirements* of the tasks composing the test must be understood by the testee through suitable instructions by the tester; and the *fundaments* of the task (i.e., the elements that it comprises) must already be familiar to the testee. Also, the testee must be motivated to perform the task. These conditions can usually be assured by the testee's demonstrating satisfactory performance on easy exemplaries of the same item types as those in the test proper.

Mental ability tests (henceforth called simply *tests*) that meet all these conditions can be made up in great variety, involving different sensory and response modalities, different media (e.g., words, numbers, symbols, pictures of familiar things, and objects), different types of task requirements (e.g., discrimination, generalization, recall, naming, comparison, decision, inference), and a wide range of task complexity. The variety of possible items and even item types seems limited only by the ingenuity of the inventors of test items.

SOME FACTS OF NATURE

When a collection of such items is given to a large representative sample of the general population under the specified standard conditions, it is found that there is an abundance of positive correlations between the items; negative correlations are very scarce and are never as large as the positive correlations, assuming, of course, that all the items are scored in such a way that what is deemed as the desirable performance on every item receives a higher score than undesirable performance. The negative correlations are not only scarce and small, they become scarcer and smaller as the number of persons increases, suggesting that the existence of negative item intercorrelations in the abilities domain is largely or entirely due to error. There is no corresponding shrinkage of the positive inter-item correlations with an increase in sample size. If a fair number of items having authentically and reliably negative correlations with the majority of items could be found, it should be possible to combine a number of such negative items to create a test that would have the usual properties of a good psychometric test in terms of internal consistency reliability and test-retest reliability. Such a test then should show large negative correlations with tests composed by sampling only from the majority of items that are positively intercorrelated. No such "negative" test has ever been created, to my knowledge. The creation of such a test is a challenge to those who doubt the phenomenon of *positive manifold,* that is, ubiquitous positive correlations among items or tests in the ability domain.

But a correlation matrix will also tend to be predominantly positive by pure mathematical necessity. While it is entirely possible (and usual) for all of the correlations among n tests to have positive values ranging between 0 and $+1$, the negative counterpart to this condition is a mathematical impossibility. In a matrix of zero-order intercorrelations, negative values are constrained. If variables A and B are negatively correlated -1, it is impossible that both can be negatively correlated with variable C, or D, or any other variable. While the average size of all the correlations in a matrix can have any positive value between 0 and $+1$, the largest possible average negative value of all the correlations in any matrix of n variables is $-1/(n - 1)$; hence, if the negative correlations are large, they must be few, and if they are not few, they must be small. Although there is a mathematical limitation on negative correlations, the proportion and size of the positive interitem correlations actually found in the ability domain far exceeds the amount of positive intercorrelations that would be expected by chance.

Yet the generally positive correlations between items, as a rule, are rather surprisingly small. Given the internal consistency reliability (K-R 20), r_{xx}, of a test of n items, the average item intercorrelation, $\bar{r}_{ij}$, is $\bar{r}_{ij} = r_{xx}/[n - r_{xx}(n - 1)]$. In the case of even such a homogeneous test as the Raven Progressive Matrices, the value of $\bar{r}_{ij}$ is only about $+.12$ or $+.13$. The small correlations are partly due to an artifact, namely, the restriction of variance as the item difficulty of dichotomously scored items departs from .50. Even after correcting for the effect of this restriction of variance on the correlations, however, it is apparent that single test items have relatively little of their variance in common. In fact, typically less than a quarter of the variance of single items overlaps the total variance of any collection of n such items, even when the items are homogeneous in type. The collection of items may be a random sample from a large pool of diverse items, in which case the average interitem correlation would be relatively low, or it may be a selection of highly similar, or homogeneous, items, in which case the average item intercorrelation will be relatively high. But even the high interitem correlations will average only something between about $+.10$ and $+.15$.

Nevertheless, interitem correlations greater than 0 and less than .15 are large enough to create a test with a very substantial proportion of reliable or true-score variance, provided the number, n, of items composing the test is large enough. This is inevitable, because the reliable variance of total scores on a test is equal to the sum of all the interitem covariances in the square matrix of interitem covariances. A test of n items with an average interitem correlation of $\bar{r}_{ij}$ will have an internal consistency (K-R 20) reliability of $r_{xx} = n\bar{r}_{ij}/[1 + (n - 1)\bar{r}_{ij}]$. Consequently, by increasing the number of items sampled from the ability domain, as previously defined, one can create a test of any desired reliability (less than 1). Most standard tests have reliabilities greater than .90 when used on samples of the general population. When a number of such highly reliable ability tests, comprising diverse contents and item types, are administered to a representative sample of the general population, the intercorrelations of the tests are all positive and generally substantial. In other words, the various tests have a lot of variance in common.

This seems to be an unavoidable fact of nature. It has proven impossible to create a number of different mental tests, each of highly homogeneous items, and with high reliability, that do not show significant correlations with one another. The "positive manifold" of test intercorrelations is indeed a reality, a fundamental fact, that calls for scientific explanation.

A hypothesized explanation of the correlation between any particular pair of different, but singly homogeneous, tests will often point to certain common surface features of the two tests that may seem to plausibly account for their correlation. But hypotheses of this kind run into greater and greater difficulty as they try to explain intercorrelations among diverse tests. The surface features of tests soon prove inadequate to the explanatory burden when the number and diversity of tests increases but still displays positive manifold. It is well-nigh impossible, for example, to account for the correlations between vocabulary, block designs, and backward digit span in terms of common features of the tests. Explanations of correlations in terms of the surface features of tests would turn out to require nearly as many explanations as there are pairs of different, but correlated, tests. From the viewpoint of scientific theory, such a multiplicity and specificity of explanations is quite unsatisfactory, if not entirely unacceptable, and, in fact, no one systematically even attempts it.

Psychometricians since Spearman have preferred to describe the intercorrelations among a number of tests in terms of a smaller number of hypothetical factors (i.e., sources of variance) that certain tests have in common. The burden of explanation, therefore, shifts from explanations of single correlations between particular pairs of tests to a much more limited number of hypothetical factors that a number of tests measure in common.

FACTOR ANALYSIS AND THE HIGHEST ORDER FACTOR

Spearman (1904, 1927) hypothesized that the positive correlation among all cognitive tests is due to a general factor that is measured by every test. His invention of factor analysis permitted estimation of the proportion of the total variance in a collection of tests that is attributable to the *general factor*, *g*, as well as the correlation (termed a *factor loading*) of each test with the *g* factor that is common to all of the tests. Variance that is not attributable to the *g* factor (call it the non-*g* variance), is assignable to (1) other factors, called *group factors,* because they account for the non-*g* correlations among only certain groups of tests, (2) *specificity,* or that portion of a test's true score (i.e., reliable) variance that is not shared in common with any other tests in the collection of tests subjected to factor analysis, and (3) *error* variance.

Aside from error variance, *specificity* is the least interesting from a psychological and psychometric standpoint, because specificity can dwindle as more tests of similar types are added to the collection; then some of the specific variance turns into additional group factors (also termed *primary,* or *first-order,* factors).

The general factor, *g,* is the highest common factor in the correlation matrix, accounting for more of the total common factor variance than any other factor, and often even more than all of the other factors combined.

A *g* factor can be extracted by any one of three methods in current use. It can be represented by (1) the first principal component of a principal components analysis, or (2) the first factor of a common factor (or principal factor) analysis, or (3) a hierarchical factor analysis, in which all of the first-order factors are rotated to an oblique "simple structure" and the correlation among the first-order factors are then factor analyzed to yield a second-order factor. The *g* factor, the apex of the hierarchy, most typically emerges as the only second-order factor, although in large and highly diverse collections of tests, *g* appears as a third-order factor at the apex of the hierarchy.

It is desirable to "residualize" the factor loadings at each level in the hierarchy, i.e., the variance that is common to the oblique (i.e., correlated) first-order factors is partialled out and transferred up to the second-order oblique factors, and their common variance also is partialled out and transferred to the third-order factor. This procedure orthogonalizes the entire hierarchy; that is, all the factors are uncorrelated with one another, within and between levels of the hierarchy. This hierarchical analysis can be accomplished by means of the Schmid-Leiman (1957) procedure, which yields the factor loadings of all the tests on each of the orthogonal factors at every level of the hierarchy. A schematic factor hierarchy is shown in Fig. 4.1.

Is there a preference among these methods of extracting a *g* factor? Yes, although each method has certain advantages and disadvantages. The first principal component is the least affected by sampling error, and the hierarchical analysis is the most affected, and therefore should be used with samples that are very much larger than the number of tests. The first principal component will always yield the largest *g* in terms of its eigenvalue or the proportion of total variance accounted for, but this is not a real advantage, because some small part of that variance consists of *uniqueness* (i.e., the specific and error variance), which is more or less evenly spread over all the components in a principal components analysis. Thus we often find that the various tests' loadings on the

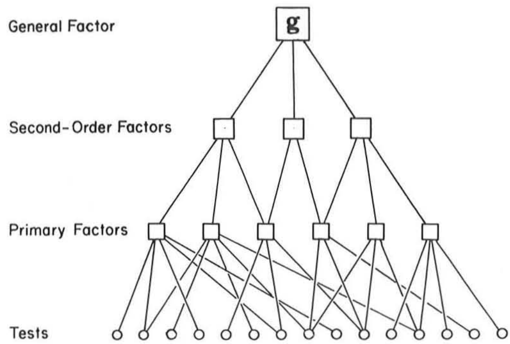

FIG. 4.1. Example of a hierarchical factor analysis with three levels.

first principal component, although they are slightly larger overall than the corresponding loadings on the first principal factor, are somewhat less clear-cut. Despite this, the first principal component and first principal factor are nearly always extremely alike. I have yet to find a correlation matrix of real tests for which the congruence coefficient between the first principal component and the first principal factor is lower than +0.99, which means that for most purposes they can be regarded as virtually identical. (This is not true of the subsequent unrotated components or factors extracted after the first; the congruence between the corresponding components and factors decreases with each successive component extracted.)

The hierarchical *g* is always smaller than the *g* represented by either the first principal component or first principal factor. This is because the process of extracting a hierarchical *g* (using the Schmid-Leiman orthogonalization transformation) does not result in any significant negative correlations in the residual matrix after the *g* factor is removed, so that positive manifold of the residual matrix is preserved when factors are partialled out at every level of the hierarchy, and virtually all of the statistically reliable factor loadings are positive on all factors. This condition is theoretically desirable in terms of thinking of all abilities as positive vectors and as always facilitating, and never hindering, performance on any cognitive task that is at all affected by the ability. (The preservation of all positive loadings on all factors was originally advocated by Thurstone (1938, 1947), as one of the aims of factor rotation to approximate simple structure.)

In extracting *g* by principal factor analysis and hierarchical factor analysis from the same set of data, I have found that the hierarchical *g* usually contains some 10% to 20% less variance than the *g* represented by the first principal factor. Yet the relative sizes of tests' loadings on the first principal factor and on the Schmid-Leiman hierarchical *g* are usually highly similar, with coefficients of congruence of +0.99 or greater. When both the first principal factor and the hierarchical *g* are extracted from the intercorrelations (based on the national standardization data) of the 13 subtests of the Wechsler Intelligence Scale for Children, for example, the coefficient of congruence between them is +0.999 (Jensen & Reynolds, 1982). I have compared both types of g factors in many collections of tests and have never found the relative magnitudes of the factor loadings to differ appreciably. However, an advantage of the hierarchical g is that it is less affected by variations in the sampling of tests entering into the analysis. For example, if we included a half-dozen or so more different types of memory span tests in the Wechsler battery, the first principal factor would be pushed somewhat in the direction of the memory factor, that is, its loadings on the memory span tests would be enlarged. The hierarchical g, however, would remain relatively unaffected by the number of tests of different types in the battery. In short, the hierarchical g is more stable than the first principal factor across variations in psychometric sampling.

When the first-order factors are rotated, the first factor loses its status as the highest common factor; its variance is scattered among the rotated primary factors, and what could properly be called a g factor disappears. The most popular rotational criterion is Thurstone's concept of *simple structure,* which aims for a factor pattern that contains no negative loadings and a maximum of zero loadings. An idealized simple structure is shown in Table 4.1. (If the factors were all orthogonal, there would be no g.) If the rotated factors are forced to be orthogonal (i.e., uncorrelated), achievement of a clean simple structure has proved to be impossible in the abilities domain. The basic assumption underlying orthogonal simple structure is that test scores are simple in factorial composition. Simple structure implies the hope that a number of tests could be devised, each of which measures only one ability, so-called primary mental abilities. But despite

TABLE 4.1
A Rotated Factor Matrix Showing Factor Loadings of an Idealized Simple Structure

	Rotated Factors				
Variable	A	B	C	D	h^2
1	1	0	0	0	1
2	1	0	0	0	1
3	1	0	0	0	1
4	0	1	0	0	1
5	0	1	0	0	1
6	0	1	0	0	1
7	0	0	1	0	1
8	0	0	1	0	1
9	0	0	0	1	1
10	0	0	0	1	1
Eigenvalue	3	3	2	2	
% Variance	30	30	20	20	

h^2 = communality

concerted efforts, this goal has never been attained. No matter how homogeneous each of a number of tests is, or how "factor pure" their constructors have striven to make them, they are always found to be substantially correlated with one another in any sizeable representative sample of the general population. When the correlations among such tests are factor analyzed and rotated to orthogonal simple structure, which is now most commonly done analytically, using Kaiser's (1958) varimax, the desired "simple structure" is never "clean," that is, instead of many near-zero factor loadings there are many low but significant loadings scattered throughout the matrix, representing the dispersal of the general factor throughout all the primary factors. Although varimax or other simple structure rotation aids in the identification and interpretation of the group factors because of the fairly sharp contrast between large and small factor loadings that serves to highlight the various primary factors, it has the disadvantage of scattering and submerging the *g* factor beyond recognition.

To overcome this problem, Thurstone suggested oblique rotation yielding correlated primary factors; this achieves a much closer approximation to simple structure. But the *g* variance then resides in the correlations among the primary factors, which, when factor analyzed, yield the *g* factor at the top of the hierarchy. Hence, in the abilities domain, it is an incomplete and unacceptable practice to stop factor extraction with orthogonal rotation of the primary factors. So, too, are oblique primary factors an incomplete analysis, unless one goes on to extract *g* (and any other higher-order factors). To pretend that *g* does not exist because it can be "rotated away" is merely deceptive. The purely mathematical argument that any position of the factor axes is as good as any other, is theoretically unacceptable. The argument rests simply on the fact that the same amount of common factor variance is accounted for regardless of the position to which the factor axes are rotated, and any factor structure (given the same number of factors) can reproduce the original zero-order correlations among the tests equally well. While it is indeed true that an unlimited number of different positions of the factor axes is possible, and that all of them are mathematically equivalent in reproducing the original correlations, some factor structures make much more sense, theoretically, than others. Some possible factor structures may even create quite misleading impressions. When we "hide" the *g* factor in the orthogonal simple-structure primary factors, for example, we create the expectation that some of the mental tests are uncorrelated, when in fact this is contradicted by the all-positive matrix of actual test of intercorrelations. Orthogonal simple structure also does not reflect the fact that the average differences *between* individuals on a number of tests are larger than the average differences between tests *within* individuals. The *g* factor, along with the smaller group factors in a hierarchical analysis, best represents all these salient facts far better than any orthogonal rotation of multiple first-order factors that dissipates *g*.

The *g* factor of a large and heterogeneous battery of mental ability tests differs in one important way from all the other rotated or unrotated factors that can be extracted, besides the fact that *g* is the single largest factor. The *g* factor cannot easily be characterized, if indeed it can be described at all, in terms of the features of the tests on which it has its most salient loadings, while all the primary factors can be characterized in terms of test content, such as verbal, numerical, spatial, and memory. When such diverse tests as Wechsler Vocabulary and Raven Matrices both have almost equally high *g* loadings when factor analyzed among a battery of diverse tests, psychological interpretations of *g* are difficult and certainly not obvious. The apparent features of the tests and the overt behavioral skills evinced by successful performance on the tests afford scant clues as to the basis for their high correlations with each other and with *g*. In attempting to characterize *g*, one is forced to seek a level of generality that transcends the "phenotypic" features of particular tests and to invoke theoretical concepts involving deeper levels of analysis. In confronting *g*, we are dealing with a highly abstract theoretical construct.

Factors, including *g*, are not themselves explanatory constructs. They are constructs which themselves require explanation. The *g* factor, above all, is a phenomenon worthy of scientific analysis and explanation. At present, we are still not very far ahead of the position noted by Spearman in 1927, when he stated that

> This general factor *g*, like all measurements anywhere, is primarily not any concrete thing but only a value or magnitude. Further, that which this magnitude measures has not been defined by declaring what it is like, but only by pointing out where it can be found. It consists in just that constituent—whatever it may be—which is common to all the abilities inter-connected by the tetrad equation. This way of indicating what *g* means is just as definite as when one indicates a card by staking on the back of it without looking at its face. Such a defining of *g* by site rather than by nature is what was meant originally when its determination was said to be only "objective." Eventually, we may or may not find reason to conclude that *g* measures something that can appropriately be called "intelligence." Such a conclusion, however, would still never be a definition of *g*, but only a "statement about it." (pp. 75–76)

I believe Spearman was quite correct in tentatively identifying intelligence only with *g* rather than with *all* of mental ability. There is no theoretical limit to the possible number of ability factors, so long as we can go on making slight variations in numerous mental tests such that their intercorrelations are less than 1 when corrected for attenuation. Hence, to equate intelligence with all of mental ability would surely render this concept scientifically undefinable and unmeasurable. If we reject this alternative, and *g* as well, as definitions of intelligence, we are left either with the problem of deciding which other factor should be included in our definition or of resorting to pure operationalism, declaring that one particular test is *the* measure of intelligence.

FLUID AND CRYSTALLIZED ABILITY

Cattell (1963, 1971) discovered that, when various tests with contents reflecting past learning experiences, cultural acquisition, and scholastic knowledge and verbal and numerical skills are factor analyzed along with tests involving novel problem solving and forms of reasoning based on analogies, series, and matrices all consisting of abstract or nonrepresentational figures, there emerges at the second level of a hierarchical analysis two factors which Cattell has labeled fluid and crystallized G, or G_f and G_c. Fluid ability, G_f, can be described as relation eduction, abstraction, and reasoning in novel problems. Crystallized ability, G_c, reflects the acquisition of specific and transferrable skills and knowledge made available by the individual's culture, education, and experience. The G_f much more nearly corresponds to Spearman's concept of *g* than does G_c. Since Cattell's hierarchical model does not go beyond the second level, Humphreys (1979) has described it as an "incomplete hierarchical model" (p. 108). Because G_f and G_c are correlated, and usually highly correlated, in an oblique solution, a substantial *g* should emerge as a third-order factor—a *g* which is essentially the same as Spearman's *g*. The degree of correlation between G_f and G_c seems to be related to a number of conditions:

1. When the persons are of similar cultural background and have had fairly equal amounts of school experience, G_f and G_c are highly correlated. In our university undergraduates, for example, the correlations between various typical tests of fluid and crystallized abilities are just about as high as the correlations between tests of the same type. And Raven's Advanced Progressive Matrices, a classical marker test for G_f, is more highly loaded (+0.80) on the overall *g* factor (first principal factor) of the Wechsler Adult Intelligence Scale than are any of the WAIS subtests themselves, even though the WAIS is generally viewed as being predominantly a test of crystallized abilities.
2. A random or representative sample of the general population shows higher correlations between G_f and G_c tests than samples with a more restricted range of ability.
3. As the collection of tests becomes larger and more varied in contents and item types, G_f and G_c become less clearly distinguishable. The total unweighted composite score on a sufficiently large and broadly representative sample of cognitive tasks is almost perfectly correlated with Spearman's *g*, that is, the highest-order *g*. Although I have not seen a definitive empirical demonstration, I venture the hypothesis that collections of tests that are considered typical measures of G_c would yield a *g* that comes increasingly closer to the *g* of a collection of tests that are considered typical measures of G_f as the number and variety of G_c-type tests increases. In other words, an increasing amount of G_f can be "distilled" out of typical G_c tests as they are sampled more broadly, because the only factor common to all the highly varied measures of crystallized abilities will

be fluid ability, G_f. The fluid aspect of G_c is increasingly siphoned into G_f, and the crystallized residue recedes into the residualized primary factors, or becomes at best merely a minor second-order factor.

Something very much like this picture is seen in two recent factor analyses of large batteries of highly varied psychometric tests selected to represent a number of the second-order factors previously identified in factor analyses by other investigators and which include G_f and G_c. When a Schmid-Leiman hierarchical factor analysis is applied to these data, G_f and G_c clearly appear as second-order factors. But when the hierarchical analysis is continued to the third level, yielding *g*, the residualized second-order G_f simply disappears; it is completely absorbed into *g*. In Gustafsson's (1984) analysis, the correlation between G_f and *g* is +1.00, and Gustafsson concludes that "the second-order factor of fluid intelligence is identical with a third-order *g*-factor" (p. 179). In this analysis, much of G_c is also "absorbed" by *g*, the correlation between them being +.76. Undheim (1981a, 198ab, 1981c) re-analyzed the correlations among the 20 tests of the Horn and Cattell (1966) study which identified G_f, G_c, and three other second-order factors (G_v—spatial visualization, G_r—fluency, and G_s—"speediness"). But Undheim carried the hierarchical analysis to the third level, yielding *g*. The residualized G_f turns out to be very small, accounting for less than half as much variance as G_c and less than one fifth as much variance as *g*. Undheim, with Gustafsson, concludes that Cattell's second-order G_f is equivalent to *g*, as defined in an orthogonalized hierarchical model—a *g* referred to by Undheim as a neo-Spearmanian *g*, because it is arrived at by a method of factor analysis quite different from Spearman's outmoded tetrad method. And the residualized G_c should not really be considered a general factor at all, but a minor second-order factor correlated with primary factors arising from tests of verbal, educational, and general cultural knowledge. G_c is practically equivalent to a residualized V:ed (verbal-educational) factor in Vernon's (1950) hierarchical model.

SIZE AND INVARIANCE OF *g*

As the first (unrotated) principal factor, *g* inevitably comprises more variance than any other factor that could be extracted from the matrix of test intercorrelations. But how large a percentage of the total variance does *g* actually account for? The answer depends on the number and diversity of the tests and the range of ability in the subject sample. To get a rough idea of the size of *g*, I have examined 20 independent correlation matrices comprising a total of more than 70 tests, such as the Wechsler battery, all the tests used in the National Longitudinal Study, the Kaufman Assessment Battery for Children, the Armed Services Vocational Aptitude Battery, the General Aptitude Test Battery, and other miscellaneous collections of tests. The tests have been administered to large and fairly representative samples of children and adults. (As all scores are age-standardized, the effects of age do not enter into the correlations.) The average percentage of variance accounted for by *g* in the 20 data sets is 42.7% (with a range from 33.4% to 61.4%). The average percentage of variance attributable to all other factors that have eigenvalues greater than 1, and thus can be said to constitute other common factors, is 15.3% (with a range from 9.6% to 22.8%)—call this the non-*g* common factor variance. The ratio of *g* variance to non-*g* common factor variance was determined for each of the 20 analyses; the mean ratio over the 20 studies is exactly 3:1; that is, *g* accounts for three times as much variance as the non-*g* common factor variance. (The *g*/non-*g* ratios ranged from 1.6 to 5.2.)

Spearman originally believed that *g* is invariant across different collections of tests, but this belief depended on the truth of his two-factor theory, namely, that the true-score variance of every test comprises only *g* variance and specific variance. But the overly simple two-factor theory had to be discarded. With the acknowledgment of group factors, the invariance of *g* across different collections of tests is no longer logically assured, but is an open empirical question. It is certainly true that the particular composition of the test battery will affect its *g*. A collection of tests in which all of them are verbal will yield a *g* which is some amalgam of both general and verbal ability and will therefore be a somewhat different *g* from a test composed of both verbal and nonverbal tests in roughly equal proportions. The degree of invariance of *g* is a function of the *number, diversity,* and cognitive *complexity* of the tests in the collection that is factor analyzed. Increasing any one or a combination of these conditions increases the similarity of the *g* factor extracted in different collections of tests.

The robustness of *g* in maintaining its identity when extracted from different test batteries, however, actually seems quite impressive. Tests with larger *g* loadings in one battery generally have large *g* loadings in most other batteries. It is a rare finding, for example, when a high-*g* test such as the Raven Matrices has a *g* loading below the median *g* in any collection of psychometric tests. When this nonverbal test is factor analyzed among just the six verbal subtests of the WAIS, for example, the size of its *g* loading is second only to that of Vocabulary. When the Raven Matrices and all 11 of the WAIS subtests, which includes five nonverbal performance tests, are factor analyzed, the Raven has the highest *g* loading among all of the tests.

Another example of the robustness of *g*: The *g* loadings of the 12 scales of the Wechsler Intelligence Scale for Children-Revised (WISC-R) were obtained for the 1868 white children in the national standardization sample. In an independent sample of 86 white children, the same 12 WISC-R subtests were factor analyzed along with the 13 subtests of the Kaufman Assessment Battery for Children (K-ABC), a mental ability test designed with the hope of being quite different from the WISC-R.[1] How similar are the WISC-R *g* loadings across two independent

[1] I am indebted to Dr. J. A. Naglieri for providing these data.

samples and when the 12 WISC-R subtests are factor analyzed as a 12 × 12 correlation matrix (the standardization sample) and as part of a 25 × 25 matrix including the 13 K-ABC subtests? The average g loadings of the WISC-R subtests in these two conditions are +0.57 and +0.58, respectively, and the rank-order correlation between the two sets of g loadings is +0.97. In short, the two g factors are practically identical, even across different samples and different collections of tests.

The robustness of g across diverse test batteries was shown long ago in a study by Garrett, Bryan, and Perl (1935), who factor analyzed a battery of six varied memory tests (meaningful prose, paired-associates, free recall of words, digit span, memory for forms, memory for objects) and extracted the g factor. This battery of tests then was factor analyzed along with four other diverse tests not especially involving memory (motor speed, vocabulary, arithmetic, form board). The g loadings of the memory tests in the two analyses were correlated .80. The overall correlation between g factor scores based on just the memory tests and g factor scores based on just the nonmemory tests was .87. This is evidence that the g of the six memory tests is very close to the g of the nonmemory tests. To be sure, the memory tests were not as highly loaded on g (average g loading = .42) as the vocabulary and arithmetic tests (average g loading = .65), but what little g the memory tests have is much the same g as found in the nonmemory tests. One would like to see larger-scale studies of this type based on many diverse psychometric tests, to determine the range of correlations between g factor scores extracted from different nonoverlapping sets of tests, controlling for reliability. My hunch is that all the g factors would be found to be highly similar.

We now have considerable evidence that g is highly consistent across different racial populations when they share the same language and general cultural background. In nine independent studies in which test batteries comprising anywhere from six to thirteen tests were administered to large representative samples of black and white Americans and a g factor was extracted separately from the correlation matrices in the black and white samples, the coefficients of congruence between the g factors obtained in the black and white samples of the nine studies ranged between +0.993 and +0.999, with a mean of +0.996. Such congruence coefficients indicate virtual identity of the g factor in the black and white populations (Jensen, 1985). (From the same data, the mean group difference in g is estimated at about 1.2 σ, where σ is the average within-group standard deviation.)

PRACTICAL EXTERNAL VALIDITY OF g

The practical predictive validity of intelligence and aptitude tests is mainly dependent on g. This has been so frequently demonstrated with respect to the prediction of scholastic achievement as to not bear further reiteration. Other factors, such as verbal and numerical factors, may enhance prediction of perfor-

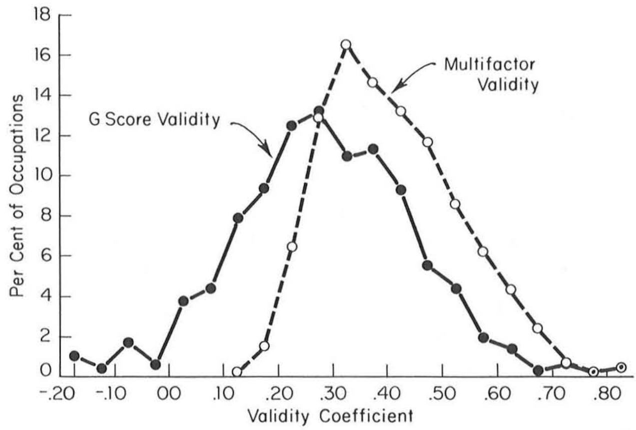

FIG. 4.2. Frequency distribution of 537 validity coefficients of the General Aptitude Test Battery for 446 different occupations. G score is general intelligence; multifactor validity is based on an optimally weighted composite of nine GATB aptitudes (including G) for each job category. The median validities are +0.27 for G and +0.36 for the multifactor composite.

mance in school and college and in the various armed forces training programs, because the predicted criterion is factorially complex, but the increases in the validity coefficient that result from adding other factors after *g* in the prediction equation are surprisingly small. The same is true for the prediction of occupational performance, although a clerical speed and accuracy factor and a spatial-visualization factor contribute significantly to the predictive validity for certain occupations. The *g* factor has predictive validity for job performance in nearly all jobs, and the validity of *g* increases with job complexity. I have found that the average predictive validities of each of the GATB aptitude tests, for 300 occupations, are substantially correlated (+.65) with the *g* loadings of these aptitude tests (Jensen, 1984). The frequency distribution of 537 GATB validity coefficients for predicting performance in 446 different jobs is shown in Fig. 4.2. The G score validity is a simple *r*, whereas the multifactor validity is a multiple *R*, which by its nature can never be less than zero and is always biased upwards. Hence, the small average difference between the two sets of validity coefficients is noteworthy. It seems very likely that no other mental ability factor or combination of factors, independent of *g*, has as many educationally, occupationally, and socially significant correlates as *g*.

THE "REALITY" OF *g*

We are frequently warned of the danger of reifying *g*, but it is never made very clear just what this might mean. Is there a danger of reifying the physicist's concept of energy, which is also an abstract theoretical construct? One and the same energy is assumed to be manifested in various forms, such as "kinetic," "chemical," and "potential" energy. Is the physicist guilty of reification when the concept of gravitation enters into his explanation of certain physical events? For nearly a century the gene was a hypothetical construct; quantitative genetics and population genetics were developed entirely in terms of this construct.

Factor analysts and intelligence theorists have always viewed g as a theoretical construct. The status of factors as theoretical constructs has been so thoroughly discussed by Burt (1940) in the chapter on "The Metaphysical Status of Factors" in his famous book *The Factors of the Mind* as to leave hardly anything more that could reasonably be said on this topic. Anyone who feels inclined to argue about this matter, I would insist, should first study Burt's masterful chapter. If it is thought that there is really anything left to argue about concerning the legitimacy of g as a bona fide theoretical construct, we should not be deprived of this enlightenment, explicated, one would hope, with the same philosophic thoroughness and scientific erudition that characterize Burt's chapter.

Recognition of g as a hypothetical construct is not to say that g represents nothing more than a mathematical artifact or a fiction entirely created by the algebraic operations of factor analysis applied to an arbitrary collection of tests. If this were proven true, g would indeed be of little scientific interest. The g factor gains interest to the extent that it is found to be significantly related to variables outside the realm of psychometric tests, from which the g construct originated. It has already been noted that a g factor dependably appears as a major hypothetical source of individual differences when we factor analyze any collection of diverse cognitive tasks on which a person's performance must meet some objectively quantifiable standard and on which task difficulty is not a function of sensory or motor skills, that is, the easy and hard tasks do not make different demands on sensorimotor abilities per se. And the g factors extracted from different collections of diverse cognitive tasks are much more highly correlated with one another than are the tasks themselves, or than are a simple unweighted sum of the scores on the tasks in each collection. Even though g is not absolutely invariant, the considerable congruence of the g factors extracted even from quite dissimilar collections of tests is consistent with the interpretation of the observed variability in g as a form of measurement error due to psychometric sampling. Variability in g arises from the fact that tests differ in their g loadings relative to other non-g factors, and most collections of tests that are submitted to factor analysis are quite limited in size. Hence there is psychometric sampling error in the g measured by any particular limited collection of tests. The resulting variability of g merely attenuates its potential correlation with external variables that might enhance its interest as a theoretical construct. In spite of such sampling variability, g is found to be related to a number of theoretically important variables which themselves have no connection whatsoever with psychometrics or factor analysis. Psychometric tests were never devised with the express purpose of predicting these variables. Here are some noteworthy examples.

Heritability of WAIS Subtests. A simple method for inferring whether there is a statistically significant proportion of genetic variance in a metric trait is Fisher's variance ratio, F, based on the within-pair variances obtained in groups of monozygotic (MZ) and dizygotic (DZ) twins; that is, $F = s^2_{WDZ}/s^2_{WMZ}$. The rationale for this ratio is that the difference between the members of a pair of DZ twins (who have, on average, only about half of their segregating genes in common) is attributable to both genetic and environmental factors, while the difference between members of a pair of MZ twins (who have identical genotypes) can be attributable only to nongenetic factors. For the genetic traits, therefore, the within-pair variance of DZ twins is necessarily greater than that of MZ twins; the F ratio reflects this difference between DZ and MZ twins, and can be used as a statistical test of its significance. An F not greater than 1 is interpreted theoretically as indicating the absence of genetic variance in the trait in question, and the more that F exceeds 1, the larger is the contribution of genetic factors to the total variance in the trait. (The precise value of $F > 1$ required for statistical significance, of course, depends on the level of significance, α, and the degrees of freedom of the numerator and denominator of the variance ratio.)

There are two independent studies in which the 11 subtests of the Wechsler Adult Intelligence Scale (WAIS) were given to samples of MZ and DZ twins and the F ratios were determined for each of the WAIS subtests (Block, 1968; Tambs, Sundet, & Magnus, 1984). (The study by Block had 60 pairs each of MZ and DZ twins; Tambs et al. had 40 pairs each of MZ and DZ twins.) The F ratios in the two studies range from 1.36 to 4.51, with a mean of 2.26; 18 out of the 22 F ratios are significant beyond the 5% level. In each study I have calculated the rank-order correlation between the profile of F ratios on the 11 WAIS subtests with the profile of g loadings of the subtests obtained from the WAIS standardization sample for ages 19 to 24 years. Thus the F ratios and g loadings are based on independent samples. The rank-order correlation between the profiles of F ratios and g loadings is $+.62$ ($p < .05$) for the Block data and $+.55$ ($p < .05$) for the Tambs et al. data. These correlations should be compared with the rank correlation of $+.62$ between the profiles of F ratios obtained in the two studies. If that correlation can be regarded as an estimate of the reliability of the F profiles, the correlation between the F and g profiles corrected for attenuation becomes $+.79$ and $+.70$, respectively. (It should be noted that test reliability itself does not enter into the F ratios, since measurement error contributes the same proportion of error variance to the within-pair differences for MZ and DZ twins alike, and the proportionality factor cancels out in the F ratio, i.e., s^2_{WDZ}/s^2_{WMZ}.) In brief, these studies show that there is a relationship between the size of g loadings of the WAIS subtests and the degree to which the subtests reflect genetic variance.

Family Correlations. Nagoshi and Johnson (1966) correlated the g loadings of 15 highly varied cognitive tests with the degree to which the tests are correlated between different pairs of family members in a large sample (927 families) of Americans of European ancestry. The correlations of the 15 tests' profile of *g* loadings with the profile of family correlations (disattenuated) on each of the 15 tests are as follows:

Between spouses	$+.90$, $p < .001$
Father-son	$+.55$, $p < .05$
Mother-son	$+.69$, $p < .01$
Father-daughter	$+.59$, $p < .05$
Mother-daughter	$+.76$, $p < .001$
Brother-brother	$+.33$
Sister-sister	$+.42$
Brother-sister	$+.26$

Nagoshi and Johnson note that the heritability of *g* (to the extent that heritability can be assessed through family correlations) appears to be higher than that of non-*g*, possibly because of greater assortative mating for *g* than for non-*g*; *g* appears to have greater influence on educational and occupational attainment than does non-*g*.

Inbreeding Depression. If the genetic factors (alleles) that enhance the phenotypic expression of a trait are dominant, the effect of inbreeding is to lower the mean of the trait in the inbred group relative to the mean of a noninbred but otherwise comparable population—a phenomenon known as "inbreeding depression." The effect depends on the presence of genetic dominance, and the presence of dominance indicates that the trait has undergone directional selection in the course of its evolution. Hence the presence of inbreeding depression, signifying dominance, in the case of psychometric tests of ability suggests that variance on such tests reflects in part a trait of biological relevance as a fitness character for which there has been positive selection in the course of human evolution.

There are now at least 12 independent studies that have reported the genetically predictable effects of inbreeding on mental test scores (reviewed by Jensen, 1983; Agrawal, Sinha, & Jensen, 1984). The effect of inbreeding depression on the IQs of the children of first-cousins, as compared with children of unrelated parents, is about one third of a standard deviation for the Wechsler IQ (Jensen, 1983) and about one half of a standard deviation on the Raven Matrices, a more purely *g*-loaded test (Agrawal et al., 1984).

The degree of inbreeding depression on the various subtests of the Wechsler Intelligence Scale for Children (WISC) is directly related to the subtests' *g* loadings. The rank-order correlation between the profile of the index of inbreeding depression on 11 WISC subtests and the profile of the subtests' *g* loadings is about +0.8 (Jensen, 1983). Varimax rotated factor loadings show markedly smaller correlations with the index of inbreeding depression than do the *g* factor loadings. These results are consistent with the hypothesis that psychometric *g* reflects to some extent a biological aspect of intelligence that acts as a fitness character which has been subjected to natural selection in the course of human evolution.

Speed of Mental Processing. A variety of reaction time (RT) tasks, or elementary cognitive tasks (ECT), have been found to be correlated with psychometric tests of intelligence and scholastic achievement (Carlson & Jensen, 1982; Carlson, Jensen, & Widaman, 1983; Carroll, 1980; Cohn, Carlson, & Jensen, 1985; Jensen, 1982a, 1982b; Jensen & Munro, 1979; Vernon, 1983; Vernon & Jensen, 1984). Not only are subjects' median RTs (measured over a number of trials) correlated with psychometric tests, but intraindividual variability (measured as the standard deviation of the subject's RTs over a number of trials) shows comparable correlations. The correlation of RT and ECTs with psychometric tests of ability seems to depend mostly, perhaps even entirely, on *g*. The remarkable thing about these simple tasks designed to measure speed of mental processing is that the tasks usually involve nothing that would ordinarily be regarded as intellectual content. The tasks are so simple and the error rates are so low that individual differences in performance usually cannot be reliably scored in terms of the number of right or wrong responses. RTs measured in milliseconds, however, when averaged over a number of test trials for each subject, yield measures with satisfactory reliability. The easiness of the tasks is suggested by median RTs that are generally less than one second.

With a sample of university students, Vernon (1983) used scores on the eleven subtests of the WAIS in a multiple regression to predict a composite RT score created by summing subjects' median reaction times and intraindividual variability after these were converted to z scores. The shrunken multiple R was substantial (.44), even in this restricted university sample (Full Scale IQ = 122, SD = 8). However, the correlation of only the g factor of the WAIS is $-.41$; that is, all the non-g variance in the 11 WAIS subtests increases the multiple R by only .03. The profile of g loadings of each of the WAIS subtests shows a rank-order correlation of $-.73$ with the profile of each of the subtests' correlations with the composite RT score, but this correlation is attenuated in this university sample which has a restricted range on g, as the lowest Full Scale IQ of any subject in the study was at the 75th percentile of the WAIS standardization sample. (The data for this analysis were provided by P. A. Vernon.)

A similar effect is seen in a study by Hemmelgarn and Kehle (1984), who used a RT apparatus like that described by Jensen and Munro (1979), in which the subject's RT to either 1, 2, 4, or 8 light-button alternatives is measured. (See Appendix for a description of this paradigm.) In this arrangement, RT is an increasing linear function of the number of bits of information in the stimulus array (i.e., bit = $\log_2 n$, where n is the number of light-button alternatives), an effect known as Hick's law. The slope of this function is regarded as a measure (inverse) of the speed of information processing, in milliseconds per bit. Hemmelgarn and Kehle correlated individual differences in the RT slope measure with scores on each of the 12 subtests of the WISC-R in a group of 59 elementary school pupils. (Chronological age was partialled out.) The profile of 12 correlations showed a rank-order correlation of $-.83$ ($p < .01$) with the profile of the subtests' *g* loadings. That is, the degree to which a WISC-R subtest is correlated with a RT index of information processing speed is related to the size of its *g* loading. The overall correlation between RT slope and Full Scale IQ was only $-.32$, but a larger correlation would hardly be expected, considering the generally low test-retest reliability of the slope measure. RT measures, and particularly the slope, are quite sensitive to physiological state, which fluctuates for individuals from day to day.

Evoked Cortical Potentials. Various parameters of the electrical potentials of the cerebral cortex evoked by visual or auditory stimuli have been found to be correlated with IQ. Haier, Robinson, Braden, and Williams (1983) conclude:

> Perhaps, the most startling conclusion suggested by this body of work is not just that there is a relationship between brain potentials and intelligence, but that the relationship is quite strong. This supports the proposition that the variance of intelligence, with all its complex manifestations, may result primarily from relatively simple differences in fundamental properties of central brain processes. (p. 598)

Eysenck and Barrett (1985) derived a measure from the average evoked potential (AEP) that reflects the *complexity* of the waveform as indicated by the contour perimeter of the AEP wave in a given time-locked epoch. Higher IQ is associated with greater complexity of the AEP waveform; correlations in excess of $+.60$ have been found between IQ and AEP. Eysenck and Barrett factor analyzed the correlations among the 11 subscales of the WAIS obtained on 219 subjects on whom there were also obtained a composite measure of AEP complexity, which subtracts the complexity measure from the variability of the AEP, as variability is negatively correlated with IQ. When the composite AEP measure was included in the factor analysis along with the 11 WAIS subtests, the AEP had a loading of $+.77$ on the *g* factor. Moreover, the profile of *g* loadings of the WAIS subtests showed a rank-order correlation of $+.95$ ($p < .01$) with the profile of correlations of each of the WAIS subtests with the AEP. (When all the correlations in each profile were corrected for attenuation, the rank-order correlation dropped to $+.93$ [$p < .01$].) In short, the *g* factor of the WAIS is shown to be highly reflected in an electrophysiological measurement of cortical activity in response to simple stimuli (auditory "clicks") that cannot be regarded as cognitive or intellectual by any conventional definition of these terms.

Following a lead from Eysenck, Schafer (1985) independently has discovered a highly similar effect based on the AEP. In a sample of 52 adults of average or superior intelligence (WAIS Full Scale IQs of 98 to 142), Schafer measured the amplitudes of AEPs to two blocks each of 25 stimuli (auditory clicks). The percentage difference between the averages of the first and second blocks was a measure of EP habituation. (Subjects show a decrease in EP amplitude over repeated trials.) This measure of EP habituation correlated +.59 ($p < .01$) with WAIS Full Scale IQ. (When corrected for the restricted range of IQ in this sample, the correlation is +.73.) A range-corrected multiple R of .80 was obtained when another index derived from the AEP was used along with the habituation measure. Schafer correlated the profile of WAIS subtest loadings on the first principal component in his sample with the profile of correlations between each of the subtests and the EP habituation index; the rank-order correlation is +.91. When the same analysis is done using the first principal factor (instead of the first principal component) to represent the *g* of the WAIS, the results are as shown in Fig. 4.3. The rank-order correlation is +.77 ($p < .01$). The *g* loadings of the WAIS subtests in Schafer's sample show a congruence coefficient of +.98 with the loadings of the same subtests in the WAIS national standardization sample and therefore can be regarded as representing the same *g*.

The idea that *g* is really no more than merely an artifact peculiar solely to conventional psychometric tests and the mathematical manipulations of factor analysis applied to the intercorrelations among tests is utterly inconsistent with these findings showing that the *g* factor, rather than other components of variance in psychometric tests, is the most highly correlated with such variables

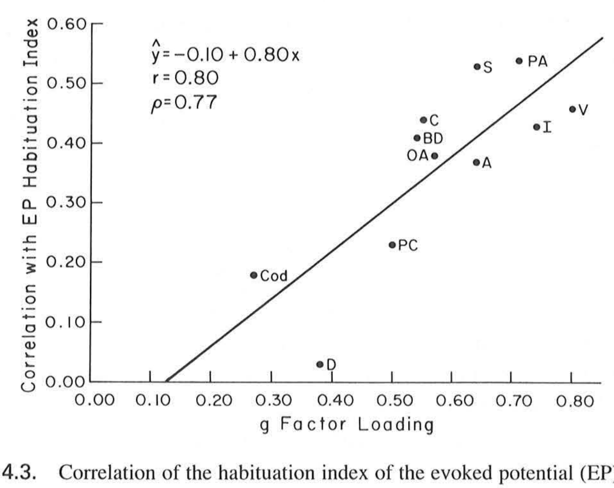

FIG. 4.3. Correlation of the habituation index of the evoked potential (EP) with Wechsler Adult Intelligence Scale (WAIS) subtests plotted as a function of the subtests' *g* loadings (i.e., first principal factor) in Schafer's study. WAIS subtests: 1—Information, 2—Comprehension, 3—Arithmetic, 4—Similarities, 5—Digit Span, 6—Vocabulary, 7—Digit Symbol, 8—Picture Completion, 9—Block Design, 10—Picture Arrangement, 11—Object Assembly.

outside the realm of psychometrics as heritability, inbreeding depression, reaction times in elementary cognitive tasks, and certain parameters of cortical evoked potentials. The alotted space does not permit a proper summary and evaluation of a number of other physical correlates of *g*, such as stature, brain size, myopia, blood types, and body chemistry. (I am presently preparing a detailed critical review of all the evidence on the physical correlates of *g*.)

The evidence reviewed here also seems to contradict the notion expressed by a modern factor analyst, Undheim (1981c), who, in criticizing the Spearman and Cattell interpretation of *g* as a ''free-floating capacity'' states that ''. . . there is no difference between intelligence and intellectual achievements. There is no measure of 'capacity,' only different measures of achievement'' (p. 257). It is hard to understand in what sense *g*-correlated reaction times and evoked potentials can be described as ''achievements'' by any generally accepted meaning of that word.

One can make various statements about *g* while not fully understanding its nature. In light of our present understanding, it would seem safe to say that *g* reflects some property or processes of the human brain that is manifested in many forms of adaptive behavior, and in which people differ, and that increases from birth to maturity, and declines in old age, and shows physiological as well as psychological or behavioral correlates, and has a hereditary component, and has been subject to natural selection as a fitness character in the course of human evolution, and has important educational, occupational, economic, and social correlates in all industrialized societies. The behavioral correlates of *g* bear a close resemblance to popular or commonsense notions of intelligence. But whether the word ''intelligence'' is attached to *g* is unimportant, scientifically. An advantage of pursuing *g* is that we have a specified set of operations on a specified class of empirical data that dependably yields a phenomenon that we can study in generally the same analytic manner that science approaches any other natural phenomenon.

REFINING *g*

The notion that *g* comes about because test constructors intentionally make up tests so that they will all be positively correlated with one another, and that they discard all tests (or test items) that are not positively correlated with all the rest, is simply false. In fact, psychometricians have often striven to devise mental tests that would *not* be correlated with one another. Thurstone (1935), for example, devoted years to trying to produce a number of tests that would yield uncorrelated measures of what he then regarded as independent factors of ability, termed primary mental abilities (PMA). No amount of psychometric refinement of the various PMA tests could eliminate their substantial intercorrelations, and, in a review of Thurstone's work, Eysenck (1939) factor analyzed all of the

Thurstone tests and found that a large *g* factor could be extracted from their intercorrelations. All but a very few of the tests had larger factor loadings on *g* than on the particular primary mental ability factors that they were specially devised to measure as purely as possible.

However, it can be argued that a correlation between two tests is not necessarily evidence that the tests measure an ability that is common to both, except in a trivial sense. That is, the common factor implied by a correlation need not be anything we could legitimately regard as an ability or a cognitive process. Common factors can arise from different causes, some more profound or intrinsic than others. If psychometric *g* could be shown to be the result of some relatively superficial common factor, it would drastically change the complexion of *g* theory. Factor analysis per se makes no assumptions about the causes of correlation and is totally indifferent to the fact that two variables may covary without sharing any common process. It could be hypothesized, for example, that *g* merely reflects cultural differences that affect a broad spectrum of cognitive skills acquisition, or nutritional differences that affect motivation and performance of all kinds. To illustrate the point in the simplest way, I can make up an analogies test on which all of my relatives will obtain much higher scores than can be obtained by any other group of people on earth. The analogies would consist entirely of items like this:

Linda is to Lydia as Leo is to: Art, Bob, Eddie, Lou. All of the names in such items are of relatives who are related as spouses, siblings, parent-child, cousins, etc. If such a test, based on the names of my relatives, were given to all my relatives and to all of yours, there would be plenty of variance, very high item intercorrelations, and a big *g* factor. This *g*, however, would have arisen entirely from the *between* families component of the correlations, and the *g* would diminish drastically, or even disappear entirely, if the correlations were obtained *within* families.

The methodology for obtaining *between*-family and *within*-family correlations among tests and for contrasting the factors extracted from the two types of correlation matrices is a way of assessing the relative proportions of wheat and chaff that we have in our *g* factor and in the *g* loadings of any given variable in the analysis. (The same can also be said in regards to any other factors.) I have explicated this methodology elsewhere (Jensen, 1980).

Does the existence of *g* depend on those sources of test score variance that differ *between* families, such as cultural and social class influences on intellectual development? If so, a *g* factor should show up only in a *between*-families factor analysis; the *g* of a *within*-families analysis should be negligible, or at least quite different. Cultural and social class sources of variance exist only *between* families. By far the larger part of what most psychologists and sociologists mean by "environment," when they speak of environmental differences that affect performance on IQ tests, refers to the between-families aspect of environmental variance. Siblings reared within the same family share the same cultural and social class influences. By factor analyzing correlations among tests between and within families, we can determine the degree to which the extracted factors are a function of between-families variance. If a factor is essentially the same both between and within families, it can be said to reflect a more intrinsic or basic source of individual differences than if it exists only between families.

Between-families (BF) and within-families (WF) correlations require a sample of *N* families, each with two or more full siblings, to each of whom are administered two or more tests on which scores are age-standardized. A BF correlation between tests X and Y, for example, is obtained by correlating the *N* family means of each set of siblings on text X with the corresponding means on test Y. A WF correlation is obtained by correlating the signed difference between siblings on test X with their difference on test Y. The WF correlation, therefore, can reflect none of the BF variance. When BF and WF correlations are obtained on a number of different tests, we can extract a *g* from each correlation matrix and compare the BF and WF *g* factors by means of the coefficient of congruence, an index of factor similarity on a scale from 0 to ± 1.

So far we have no really ideal study of this type in terms of a sufficiently broad sample of tests. But three independent large sets of sibling data that I have analyzed give such consistent results as to suggest that other collections of cognitive ability tests would probably lead to the same conclusion. In one study (Jensen, 1980), children in 1,495 white families and 901 black families in grades 2 to 6 were given seven tests: memory, figure copying, pictorial IQ, nonverbal reasoning (figure analogies, matrices), verbal IQ, vocabulary, and reading comprehension. Only the two siblings most similar in age in each family were used. BF and WF intercorrelations of the tests were factor analyzed separately for black and white samples. The coefficients of congruence between the BF *g* and the WF *g* were +.985 and +.987 for the black and white samples, respectively. In other words, the *g* factors extracted from the BF and WF correlations are practically identical in this collection of tests, for both black and white children. (The average congruence coefficient between the black and white *g* factors is +.991.)

In an independent study, being prepared for publication, four of Thurstone's Primary Mental Ability tests (Verbal, Numerical, Spatial, and Reasoning) and Cattell's Test of *g* (from Cattell's 16 P.F. battery) were obtained on 313 siblings in 135 white families. The coefficient of congruence between the BF *g* and WF *g* is +.98.

It has been hypothesized that the intercorrelation of otherwise uncorrelated abilities, thereby giving rise to *g*, comes about as a result of cross-assortative mating for various abilities (Price, 1936). If each of two abilities is influenced by entirely separate sets of genes, and if both abilities are socially perceived as desirable, there will tend to be cross-assortative mating for the abilities. That is, not only will like attract like for either ability alone, but the separate abilities will be perceived with some degree of equivalence in terms of desirability, and there will be a marital correlation between the two abilities. This common assortment of the genes that affect two traits results in a genetic correlation between the traits in the offspring—but it is only a *between*-families genetic correlation. Because the separate genes segregate in the process of gametogenesis and each offspring of a given pair of parents receives a random half of each parent's genes, there will be no *within*-family genetic correlation between the traits that are genetically correlated in the population.

Hence a test of the hypothesis that *g* arises from genetic correlations due to cross-assortative mating for otherwise genetically independent abilities consists of a comparison of the BF and WF correlations between measures of different abilities.

The correlation of about +.2 between height and IQ appears to be this type of adventitious genetic correlation due to cross-assortative mating for stature and intelligence. Although the population correlation between height and IQ is a quite reliable phenomenon, no correlation has been found *within* families. Gifted children, for example, are taller than their nongifted age peers in the population, but they are not taller than their nongifted siblings.

A within-family genetic correlation between traits is usually attributable to *pleiotropy,* that is, the same gene affects two or more phenotypically distinct traits.

So far there have been too few studies of the genetic basis of correlated traits to permit any compelling conclusions. The results of the two BF and WF factor analyses previously mentioned, however, suggest that the correlations between abilities are probably not explainable in terms of cross-assortative mating for different abilities. But a satisfactory answer must await more detailed and systematic BF and WF correlational studies that are specifically designed to answer this question. The outcome of studies based on WF factor analysis has extremely important implications not only for the theory of *g,* but for the structural representation of all the abilities identified by factor analysis. The same method can be applied to chronometric measurements of processing components.

If there is any hope at all for identifying independent or uncorrelated elementary cognitive processes, it will be realized in the study of WF correlations. The study of abilities, throughout most of its history, has shown an obsession with independence. Many theorists have pursued it, hoping to discover components of ability that are truly independent in a more real sense than part of the uncorrelated residual variance of two (or more) ability tests after their common factor is partialled out. The desire for real components that are uncorrelated has been the philosopher's stone of psychometrics; it seems to be a philosophic position, not one dictated by scientific necessity. Since psychologists have not succeeded in devising psychometric tests that are uncorrelated, the search for this presumably desirable condition has moved on to the measurement of elementary cognitive processes. By measuring smaller and smaller components of performance on cognitive tasks, presumably, correlations between them, and hence *g,* will vanish. But it might well turn out that positive correlations between any measurable components of ability will vanish only at the point where correlation becomes impossible, that is, where there is no true variance in one (or both) of the correlated components. Just where in a reductionist analysis that point will be found we cannot say at present, but it is not impossible that variance and intercorrelations could be found all the way down to the level of neural structure and biochemical activity, just short of the molecules, or even atoms, that compose the brain. The well established substantial heritability of individual differences in *g* indicates that there is some biological substrate of individual difference in *g*, presumably in the neural structure and physiology of the cerebral cortex.

TASK COMPLEXITY AND *g*

Probably the most undisputed fact about *g* is that the *g* loadings of cognitive tasks are an increasing monotonic function of the perceived complexity of the tasks. Subjective judgments of task complexity are a fairly accurate predictor of the rank order of the tasks' *g* loadings. In general, *g* loadings decrease monotonically for tasks classified as relational, associative, perceptual, and sensorimotor. An especially clear demonstration of this is a factor analytic study by Maxwell (1972), who regards the relationship between *g* and task complexity as highly consistent with Thomson's (1948) sampling theory of *g*, which posits overlapping neural elements or bonds sampled by different tests. More complex tests presumably sample a larger proportion of the total available elements and therefore would have a greater amount of overlap than relatively simple tasks. But Spearman's theory of *g* as a general mental energy that is available for any cognitive task is equally consistent with Maxwell's results. Successful performance on the more complex tasks simply requires more mental energy. Spearman characterized *g* as "the eduction of relations and correlates" on the basis of his finding that tests involving relation eduction consistently had the largest *g* loadings of any of the many types of tests that he included in his factor analyses.

The fact that much simpler tasks than those involving relation eduction, even tasks that do not require any kind of reasoning at all, are also *g* loaded, albeit to a lesser degree, indicates that Spearman's own characterization of *g* is much too limited.

The apparent failure of the Galton and Cattell attempts to measure intelligence with quite simple "brass instrument" laboratory tests, such as various types of sensory discrimination and reaction time, and Binet's success, using much more complex tasks, led to the strongly entrenched belief among psychologists that complex tasks are an essential condition for the measurement of intelligence. Yet if intelligence tests are distinguished by very high *g* loadings, it is then also true that they differ from the much simpler tasks of the Galton-Cattell variety only in degree, for tests' *g* loadings vary in a perfectly continuous manner, ranging from values close to 1.00 on down to near 0.

A high level of task complexity, therefore, appears to be a sufficient but not necessary condition for the emergence of *g*. Some significant, positive, nonzero *g* loading is evident even in simple sensory discrimination tasks and simple reaction time (RT). As these simple tasks are made slightly more complex, their *g* loadings increase. Choice RT is more *g* loaded than simple RT, dual sensory discrimination tasks are more *g* loaded than single discrimination, and backward digit span is more *g* loaded than forward digit span. Various elementary cognitive tasks (ECTs) can be rank ordered in degree of complexity on the basis of the mean response latencies in performing the tasks. The rank order is highly correlated with the rank order of the tasks' correlations with psychometric *g* derived from unspeeded complex tests of reasoning and general knowledge. The ECTs here referred to are so simple that their mean response latencies are less than 1.5 seconds for average adults. Yet even these simple tasks are *g* loaded, and the loadings increase with task complexity as indexed by mean latency. Figure 4.4 shows the correlation of each of eight very simple ECTs with *g* factor scores derived from the ten subtests of the Armed Services Vocational Aptitude Battery (ASVAB). (The tasks are described in Jensen, 1985, p. 209.)

It will be noticed in Fig. 4.4 that the correlations of the single ECTs with the ASVAB *g* scores are all quite low, ranging from less than +.10 to about +.35.

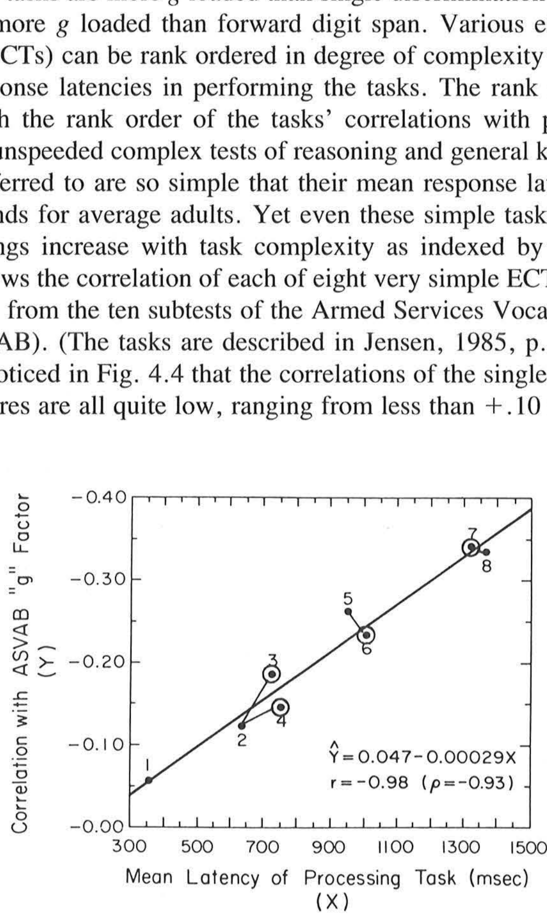

FIG. 4.4. Correlation of ECTs with ASVAB *g* factor scores as a function of task complexity as indexed by mean response latency (RT in msec.) on each task in a vocational college sample (N = 106). The dual tasks (#3, 4, 6, 7) are shown as circled dots and are connected to their single-task counterparts (#2, 5, 8) by straight lines. The tasks are described in the Appendix. (The numbers beside the data points indicate the specific processing tasks: 1—RT, 2—DIGIT, 3—DT2 Digits, 4—DT3 Digits, 5—SD2, 6—DT2 Words, 7—DT3 Words, 8—SA2.)

The shrunken multiple *R* between all eight of the ECTs and the ASVAB *g*, however, is .47, which can be compared with the average of the correlations among the ten ASVAB subtests: $\bar{r} = +.36$, $SD = .19$.

Findings such as this raise the interesting question of whether all of the *g* variance derived from very complex psychometric tests of reasoning, problem solving, and the like, can possibly be predicted by a composite score on a sufficient number of elementary cognitive tasks, none of which involves more than a very simple level of complexity. Another way of asking the same question: Is there nothing in *g* that depends upon the higher mental processes, or the so-called *metaprocesses?*

This is one of the key questions in this field, and it has not yet been adequately investigated. It is not enough to use just a few simple tasks, however reliable the scores may be made by repeated measurements. By *simple* tasks I mean ECTs that provide chronometric data such as choice RT in the Hick paradigm, speed of scanning short-term memory in the S. Sternberg paradigm, and speed of access to overlearned verbal codes in long-term memory as in the Posner paradigm. (I have described these paradigms elsewhere [Jensen, 1982a].) Each such task is much like a very homogeneous psychometric test in which all the items are of the same type. Most such homogeneous tests have a great deal of specificity (i.e., task-specific variance) and consequently not much *g* or other common-factor variance. Yet these ECTs are positively correlated with one another, and each is also correlated with the *g* factor of psychometric tests. But these single-task correlations are generally quite low, mostly falling between .3 and .4 in unrestricted samples, and even with proper corrections for attenuation, the upper limit of correlation is not greater than .50. A composite score derived from several different ECTs, however, can show larger correlations with psychometric *g*, because the total variance of a composite reflects the covariances among the components more than the variances that are specific to each component, and the covariances contain the *g* of the ECTs, some part of which is the same as the *g* of psychometric tests. It seems a likely possibility that if response latencies on as many as a dozen or so simple but distinctly different chronometric ECT paradigms were optimally combined, the composite score would correlate about as highly with psychometric *g* as do, say, the Raven Matrices, or Cattell's Culture-Fair Test of *g*, or the Wechsler, or the Stanford-Binet. Yet none of the ECTs entering into the composite score would involve anything that would ordinarily be regarded as intellectual content or as requiring reasoning or problem solving in the generally accepted sense of these terms.

Although correlations of the magnitudes being found between single ECTs and single psychometric tests may seem rather small, they should not be cause for despair. Remember that chronometric ECTs have virtually no method variance in common with unspeeded psychometric tests. It is instructive to compare the typical .3 to .4 correlations between ECTs and psychometric tests with the correlations between various psychometric tests in terms of each of their com-

TABLE 4.2
Components of Correlations[a] Among Subtests of the WISC-R Derived from Factor Loadings in a Schmid-Leiman Orthogonal Hierarchical Factor Analysis, with *g* Correlations Below the Diagonal, and Correlations Based on the Group Factors (Verbal, Memory, and Performance) Above the Diagonal

WISC-R Subtest	I	S	V	C	A	DS	TS	Cod	PC	PA	BD	OA	M
Information		13	16	13									
Similarities	45		17	13	verbal								
Vocabulary	48	48		17									
Comprehension	40	40	43				memory						
Arithmetic	38	38	41	34		16	14	08					
Digit span	29	29	32	26	25		21	12					
Tapping span	23	23	25	21	20	15		11			performance		
Coding	25	25	27	22	21	16	13						
Picture completion	34	34	37	31	29	22	18	19		08	15	15	09
Picture arrangement	33	33	35	29	28	21	17	18	25		12	12	07
Block design	43	43	47	39	37	29	23	24	33	32		22	14
Object assembly	33	33	36	30	29	22	17	19	25	25	33		14
Mazes	25	25	27	22	21	16	13	14	19	18	24	19	

[a] Decimals omitted.

mon factors. Table 4.2 shows the factor-generated correlations among the WISC-R subtests in the white standardization sample, representing the full range of ability in the white population. Below the diagonal are the correlations due to the *g* factor, in a Schmid-Leiman hierarchical analysis. Above the diagonal are the correlations among tests due to the group factors, Verbal, Memory, and Performance, orthogonal to *g* and to one another. (Correlations not significantly greater than zero at the .05 level, with $N = 1868$, are not included.) If ECTs are correlated only with the *g* factor of psychometric tests, we should expect the correlations to fall in the same ballpark as the correlations among psychometric tests that are due entirely to *g*. Such correlations, shown below the diagonal in Table 4.2, range from +.13 to +.48, with a mean of +.28.

Experimental Manipulation of Complexity. The *g* loadings of tests may be related to their complexity because responses to test items are scored as pass or fail (i.e., "right" or "wrong") and individuals' scores are determined by the threshold on the continuum of item difficulty at which the information processing system is inadequate to the task. The efficiency or capacity of the processing system may be revealed most clearly when the system is pushed or strained. Individual differences in the threshold of breakdown of the system may provide the most efficient measure of *g*

The processing difficulty of an item can be measured in terms of percent failing the item, if it is difficult enough to allow failure, or in terms of mean response latency when the item is easy enough for subjects to pass it. This hypothesis was tested in an extreme fashion by one of my graduate students (Paul, 1984). The Semantic Verification Test (SVT) consists of 14 item types, or conditions, each presented six times with different permutations of the three letters ABC. The 14 conditions are shown in Table 4.3. Following each item is

TABLE 4.3
The Fourteen Conditions of the Semantic Verification Test

SVT Variable	Semantic Condition
1	___before___
2	___not before___
3	___after___
4	___not after___
5	___first___
6	___not first___
7	___last___
8	___not last___
9	___between___&___
10	___not between___&___
11	___before___&___
12	___not before___&___
13	___after___&___
14	___not after___&___

some permutation of ABC which either agrees ("true") with the preceding statement or disagrees with it ("false"). The subject responds True or False to each item. When the SVT is given as a chronometric task to university students, the correlation between their median RTs and scores on the untimed Advanced Raven Matrices test is about −.50. Considering the great simplicity and lack of intellectual content of the SVT, and the restricted range of ability in the university group, this is a remarkably high correlation. A high level test of verbal knowledge and reasoning, Terman's Concept Mastery Test, is correlated about +.50 with the Advanced Raven Matrices in the university population, and WAIS Vocabulary, the most highly *g* loaded of the 12 WAIS subtests, is correlated only +.44 with the Raven.

The SVT was given as an untimed paper-and-pencil test to 77 third-grade pupils to determine the percent failing each item. The SVT test was also given as a chronometric task to 50 university students. The mean median RTs to the 14 conditions of the SVT ranged from about 650 msec to 1200 msec, and the overall error rate was 7%. The task was obviously of trivial difficulty for university students. The interesting point, however, is that the difficulty levels (percent failure) of the 14 conditions for the third graders shows a rank-order correlation of +.79 (disattenuated = +.83) with the mean median RTs of the 14 SVT conditions in the university sample. In university students taking the SVT as a chronometric test, the correlation of mean error rates on the 14 SVT conditions with the corresponding mean median RTs was +.82. Twenty-five university students were also asked to rank the 14 SVT conditions in the order of their complexity, according to the students' subjective judgments of complexity. The average correlation between subjects' rankings was +.80 and the reliability of the composite rank order of the 25 complexity rankings was +.99. This judged complexity of each of the 14 SVT conditions was correlated +.86 with the difficulty levels of the 14 conditions in the third graders and +.82 with the mean median RTs of the university sample. Hence there is a close relationship between judged item complexity, item difficulty (measured as percent failing), and item processing times.

These SVT RT data, however, present a seeming paradox with respect to psychometric *g* as measured by the Advanced Raven Matrices. Although the correlations between Raven scores and the median RTs of the 14 SVT conditions range between −.30 and −.50, the degree of correlation is *inversely* related to task complexity as indicated by median RT or judged complexity. The correlation between tasks' median RT and their correlation with the Raven is −.67, that is, the *less* complex SVT conditions show the *higher* correlation with Raven scores. Another paradox: although the *positive* SVT conditions (e.g., A before B) are less complex and have RTs that average 210 msec less than the *negative* SVT conditions (e.g., A not before B), the mean correlation of the RT for positive SVT items with the Raven is −.42, as compared with −.39 for the negative items (disattenuated, these are −.45 and −.43, respectively). And when the RTs of the 14 SVT items are factor analyzed, the positive items have the higher mean loading on the first principal factor (.91 vs. .88; disattenuated, .99 vs. .96). It had been hypothesized that the negative condition would necessitate an extra mental manipulation in the processing to produce a correct response and that this increased complexity would increase the item's *g* and its correlation with the Raven. Although the negative items are clearly judged as being more complex and have longer RTs (by 210 msec, on average), they are not more highly correlated with a marker test of psychometric *g*. It is surprising and puzzling. We plan to repeat the study to see if this paradoxical result is replicated.

Another experimental manipulation of complexity is by means of comparing RTs to *single* and *dual* tasks. If tasks A and B are performed separately in such a way that performance on one does not affect performance on the other, they are termed *single* tasks. If they are presented simultaneously or in close temporal proximity in such a way that performance on either A or B is significantly affected by their proximity, then the task on which performance is measured (usually chronometrically) is termed a *dual* task. (Dual tasks are also referred to as *competing* tasks.) The effect of dual tasks is commonly interpreted as dividing attention and straining processing capacity. The effect of this generally is to increase the *g* loading of the dual task relative to its *g* loading as a single task. In a dichotic listening task, for example, the subject simultaneously hears a different pattern of three notes in each ear (e.g., left ear: high, low, high; right ear: low, high, low) and is then randomly postcued to report the pattern presented to one ear. Using such paradigms, Stankov (1983; also see Fogarty & Stankov, 1982) discovered that performances are more highly intercorrelated and therefore more *g* loaded when presented as dual than as single tasks. Dual tasks were also more highly correlated with subjects' educational level than their single-task counterparts. In the most thorough study of a wide variety of dual tasks that I have come across in the literature, Fogarty (1984) found that dual tasks have higher *g* loadings than their single-task counterparts only when the latter have relatively low *g* loadings. Tasks that have high *g* loadings when presented as single tasks, however, have somewhat lower *g* loadings when they are presented as a dual task. Presumably, when a task is already high *g* as a single task, making it a dual task strains processing capacity to the point of breakdown, which lowers the reliability of the performance by increasing the rate of chance successes and consequently attenuates the task's *g* loading. Fogarty's factor analysis of single and dual tasks also suggests, although not very strongly, that dual tasks are factorially more complex than the single component tasks and that dual tasks may involve cognitive processes that are not operative in single tasks. But the evidence for this is weak and ambiguous, and in a study explicitly addressed to this question, Lansman, Poltrock, and Hunt (1983) found no evidence for any distinct abilities to divide or focus attention.

The importance of the relationship between single vs. dual tasks and *g* is that the increase in *g* loading must be purely a process phenomenon arising from the greater strain placed on cognitive capacity by dual tasks. There is no increase in the informational content of the dual task.

In our own lab we have worked with four single and dual tasks (Jensen, 1985; Vernon, 1983; Vernon & Jensen, 1984). Our various ECTs, in which performance is always measured in terms of median RT, are described in the Appendix (taken from Jensen, 1985, p. 209). Returning to Fig. 4.4, which shows the relationship between task complexity (as indicated by the mean latency, or RT, on the task) and the task's correlation with the *g* factor scores derived from the Armed Services Vocational Aptitude Battery (ASVAB) in a sample of 106 vocational college students, we see that the correlation between these variables is quite large, $r = -.98$, $\rho = -.93$. It appears anomalous, however, that one of the four dual tasks (#6) has a slightly lesser correlation with *g* than its single-task counterpart. These correlations are so similar, however, that this reversal might be due to sampling or measurement error. Another way of looking at this relationship is in terms of mean differences in median RTs between two groups that differ in general ability, or *g*. The mean differences between two contrasting groups should be less attenuated by measurement error. Figure 4.5 shows the correlation between the complexity of the processing tasks, as indicated by their mean latency (RT), and the mean difference between vocational college students and university students; both groups are normal youths of comparable age, and both groups are of above-average intelligence, although they differ about one standard deviation in psychometric *g*. As seen in Fig. 4.5, there is a high correlation ($r = +.97$, $\rho = +.98$) between task complexity and the degree to which the tests discriminate between the vocational and university groups. Also, in every case, the dual tasks show greater discrimination than their single-task counterparts. These data are highly consistent with the hypothesis that dual tasks, or task competition, increases *g* loading.

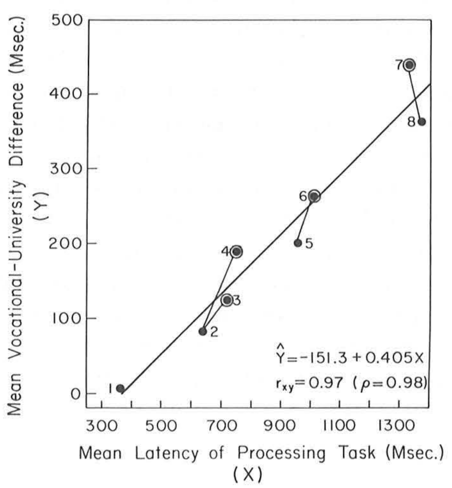

FIG. 4.5. Mean difference (in msec.) between vocational college students ($N = 106$) and university students ($N = 100$) on various elementary cognitive tasks as a function of task complexity as indicated by mean response latency (RT) on each of the tasks in the vocational college group. The same task, when presented as part of a dual task, is shown as a circled dot connected to its single-task counterpart by a straight line. Note that in every case, the dual tasks are more discriminating between the vocational and university groups than the single tasks. The tasks are the same as those in Fig. 4.4.

THEORIES OF *g*

Ever since Galton originally propounded the notion of intelligence as a general ability which could be channelled into any kind of intellectual activity, and Binet advanced the idea of intelligence as the average level of a number of different abilities and skills, various theories of intelligence, and of *g*, have been classifiable into two broad categories: *unitary* theories and *multiple* theories. The same divisions might also be labeled *power* theories and *sampling* theories, respectively. This division of theoretical conceptions has continued down to the present day. One of the major challenges to the field at present is to achieve a satisfactory theoretical resolution and consensus on the problem of the unitary or multiple nature of *g* based on empirical evidence. The answer may depend on the level of analysis we choose for our study of cognitive abilities. In formulating laws of mechanics, matter can be regarded as unitary—the solid, seeable, touchable, solid objects in our surroundings. For most of the laws of chemistry, matter is seen as multiple at the level of mixtures, compounds, and molecules, but as unitary at the level of atoms. In subatomic physics, atoms are no longer unitary but are seen as composed of multiple particles—protons, neutrons, etc., which are also analyzable into more elemental components, the quarks, and there is still no assurance that even the quarks are the ultimate units of matter that defy further analysis.

Unitary Theories of *g*

Spearman's "Mental Energy". Spearman suggested that *g* is a "mental energy" of which there is a limited amount for each individual and in which individuals differ. The brain's "energy" can be directed to any kind of mental activity executed by different "neural machines." Individual differences in the "machines" show up as group factors and, along with their complex interactions, as specificity. The overall positive correlations among these activities is all being powered by the same general energy, in which individuals differ. To quote Spearman's (1923/1973) own most succinct and explicit statement of this theory: "The brain may be regarded (pending further information) as able to switch the bulk of its energy from any one group to any other group of neurons; as before, accordingly, the amount and the direction of the disposable energy regulate respectively the intensity and the quality of the ensuing mental process" (p. 346). Elsewhere he elaborates: "In this manner, successful action would always depend, partly on the potential energy developed in the whole cortex, and partly on the efficiency of the specific group of neurons involved. The relative influences of these two factors could vary greatly according to the kind of operation; some kinds would depend more on the potential of the energy, others more on the efficiency of the engine" (1923/1973, p. 6).

I have used the word "energy" in quotes in this context, because it is not always clear whether Spearman endows the term with the meaning it has in the physical sciences, which is its only scientifically legitimate meaning, or whether he intends it merely as an analogy or metaphor. If *g* is equated with energy in the accepted physical sense of the term, then, as Thomson (1948, p. 58) pointed out, Spearman's theory can be rejected in its literal form, because the brain (or the cerebral cortex) does not act as a reservoir of free-floating energy that can be consolidated and shifted around from one group of neurons to another. Whatever energy exists in the brain resides within the individual nerve cells as an electrochemical reaction propogated along the neural membrane. If, on the other hand, Spearman's use of "energy" is merely metaphorical, it contributes little, if anything, to the scientific understanding of *g*. It merely underscores Spearman's belief in the unitary nature of the cause of *g* but does not suggest what this unitary cause is in empirically testable terms. Spearman's "mental energy" theory of *g* has always been regarded metaphorically by most psychologists, and consequently has not been taken very seriously. As metaphor, it has been peculiarly unfruitful in generating empirical investigation, and today Spearman's "energy" theory has only the status of an historical relic.

Burt's Neurophysiological Theory. Burt (1940, p. 217; 1961) proposed a unitary theory of *g* that is not metaphoric, but anatomical and physiological. He held that *g* reflects the general character of the individual's brain tissue, such as the degree of systematic complexity and organization in the neural architecture, and he cites histological evidence that the cerebral cortex of some mentally deficient persons shows less density and branching of neurons than is seen in the brains of normal persons. To account for the ubiquity of *g*, Burt hypothesizes that the general quality of an individual's cerebral cortex is more or less homogeneous throughout; hence every intellectual function would reflect this homogeneous quality of the nervous system. As with Spearman's theory, specialized areas or neural structures, in addition to particular classes of acquired knowledge and skills, give rise to group factors and specificity. Burt's theory, being nonmetaphoric, has the virtue of being testable, at least in principle, but I am not aware that, so far, there have been any systematic histological investigations of individual differences in the brain's architectonics in relation to psychometric *g* among normal persons. There is little that psychologists as such can do to confirm or substantiate Burt's theory, and so it has attracted little attention.

Motivation or Drive Theories of g. A number of Spearman's contemporaries, such as Maxwell Garnett, suggested that *g* results from individual differences in will, motivation, or drive level, which affects performance on all cognitive tasks (see Spearman, 1927, pp. 88–89). Essentially the same notion has been recently revived by Macphail (1985), who equates *g* with Hull's *D* (for *d*rive). This theory runs into difficulty on at least three grounds.

First, no independent evidence has been brought forth to show that high-*g* persons are more highly motivated in test-taking situations than low-*g* persons.

Differences in range and intensity of intellectual interests are more likely a result than a cause of differences in g.

Second, a theory of g as D runs into trouble with the Yerkes-Dodson law, the empirical generalization that the optimal drive level for error-free or efficient performance of a task is lower for simple than for complex tasks. Yet cognitively complex tasks are generally more g loaded than simple tasks, and high- and low-g individuals differ more on complex than on simple tasks. We should predict just the opposite if g were equated with D. (No one has yet proposed an inverse equation of g with D.)

Third, there is direct empirical evidence showing that higher levels of ability in a cognitive task are not associated with higher motivation or arousal during task performance, as measured independently by pupillary dilation, a sensitive indicator of motivational arousal and effort. Ahern and Beatty (1979) measured the degree of pupillary dilation as an indicator of effort and autonomic arousal when subjects are presented with test problems. They found that (1) pupillary dilation is directly related to level of problem difficulty (as indexed both by the objective complexity of the problem and the percentage of subjects giving the correct answer), and (2) subjects with higher psychometrically measured intelligence show less pupillary dilation to problems at any given level of difficulty. (All subjects were university students.) Ahern and Beatty concluded:

> These results help to clarify the biological basis of psychometrically-defined intelligence. They suggest that more intelligent individuals do not solve a tractable cognitive problem by bringing increased activation, ''mental energy'' or ''mental effort'' to bear. On the contrary, these individuals show less task-induced activation in solving a problem of a given level of difficulty. This suggests that individuals differing in intelligence must also differ in the efficiency of those brain processes which mediate the particular cognitive task. (p. 1292)

Speed of Processing and Neuronal Errors in Transmission as the Basis of g. Unitary theories of g necessarily hypothesize individual differences in some extremely basic attribute that plausibly could affect every kind of cognitive performance. Galton originally hypothesized mental speed, and proposed using RT to visual and auditory stimuli as a measure of general ability.

Galton's own efforts and those of his leading American disciple, James McKeen Cattell, were notably unsuccessful in establishing any substantial relationship between RT and independent criteria of intellectual ability, and the pursuit of intellectual correlates of RT was virtually abandoned for more than half a century.

In the past decade, however, with the development of relatively sophisticated chronometric techniques in experimental cognitive psychology (e.g., Posner, 1978), this line of research has been vigorously pursued by many investigators. As a result, many different g-loaded psychometric tests have been found to show significant correlations with RT measurements derived from a considerable variety of cognitive tasks ranging in complexity from simple RT (response to the onset of a single stimulus) to response latencies in verbal and figural analogies. I have reviewed the research on many of these RT tasks and their relationship to psychometric *g* elsewhere (Jensen, 1982a, 1982b).

Correlations between RTs measured in different paradigms are highly positive, indicating a large general speed factor that loads in a wide variety of ECTs. This general speed factor is correlated with the psychometric *g* derived from nonspeeded traditional tests of intelligence, both verbal and nonverbal.

The correlation between psychometric *g* and speed on ECTs increases with the complexity of the ECT only up to a point; beyond it the correlation diminishes with increasing task complexity. The reason is probably that the more complex tasks invite different strategies for attaining the preferred response and these tend to confound individual differences in sheer speed of mental processing with individual differences in choice of strategy. In the great variety of psychometric test items, on the other hand, strategy effects become relegated to specificity or narrow group factors, and the *g* factor reflects the more fundamental attribute of mental speed. Hence psychometric *g* is more highly correlated with relatively simple ECTs that do not invite a variety of solution strategies.

Not only speed is correlated with *g*, but also the consistency of RTs to the same task over repeated trials. We measure intraindividual variability in RT in terms of the standard deviation of RT over *n* trials, signified as σ_i. This measure is often more highly correlated (negatively) with psychometric *g* than is the median RT, despite the usually higher reliability of the median RT.

Mean differences in these parameters between criterion groups selected from different regions of the IQ distribution have shown more consistent and clear-cut results than correlations between these parameters and psychometric test scores within groups. The reason for this seems to be that correlations are always attenuated by unreliability of measurement and restriction of the range of ability, whereas a mean group difference is little affected by these factors. Differences between clearly separated criterion groups are more capable than correlations of detecting the more subtle effects in various RT paradigms.

One of our recent studies (Cohn, Carlson, & Jensen, 1985) illustrates the contrasts in mental speed between academically gifted and nongifted youths (ages 12 to 14 years) on a variety of ECTs (described in the Appendix) ranging in complexity from simple and choice RTs, to S. Sternberg's short-term memory scan for digits, to discriminating physically same vs. different word pairs, and discriminating simple synonyms vs. antonyms. All but the simple and choice RT tasks were presented both as single and as dual tasks (DT). The gifted (G) group ($N = 60$), with an average age of 13.5 years, consisted of manifestly talented youths whose scores on the SAT were on a par with university students five to six years older. The G subjects were enrolled in university courses, competing successfully in a predominantly math and science curriculum. The nongifted

(NG) group consisted of 70 white junior high school students averaging about 1 *SD* above statewide norms on the California Test of Basic Abilities. The G and NG groups differed 1.9 *SD* on the Raven Standard Progressive Matrices. For both the G and NG groups, the chronometric tasks were of trivial difficulty, with mean response latencies never as long as 2 seconds, even in the NG group.

Figure 4.6 shows the mean latencies on the eight mental processing tasks for the G and NG groups and a group of 50 U.C., Berkeley undergraduates (Un). The rank-order correlations between the shapes of the profiles are all +.98 or above. Groups G and NG differ significantly ($p < .01$ to .001) on all of the tasks, but G and Un show no significant differences. (G and Un differ only a nonsignificant 2 points on the Raven Matrices.) The within-group multiple correlation of the eight processing tasks with Raven Matrices is .60 and .50 for groups G and NG, respectively.

Most remarkable is the difference between the G and NG groups on the Hick paradigm, since it has the least intellectual content of any of the tasks, requiring only that the subject release a pushbutton when a light goes on among an array of either 1, 2, 4, or 8 lights (corresponding to 0, 1, 2, and 3 bits of information). Figure 4.7 shows the results. The groups differ beyond the .001 level at every level of task complexity from 0 to 3 bits, for both RT and MT (the interval between releasing the home button and pressing the button adjacent to the light). Also, the slopes of RT for the G and NG groups differ by .70 *SD*s, which is highly significant ($p < .001$), and intraindividual variability in RT differs significantly at every level of bits.

Such findings show that psychometric *g* can be measured by means of tests that have little or no knowledge content and that require no complex problem-solving strategies. In these respects, they are very unlike ordinary IQ tests, yet

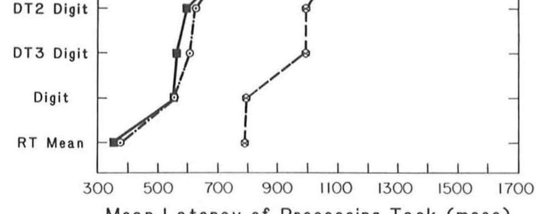

FIG. 4.6. Mean latency of various processing tasks in three groups: university students (Un), gifted (G), and nongifted (NG).

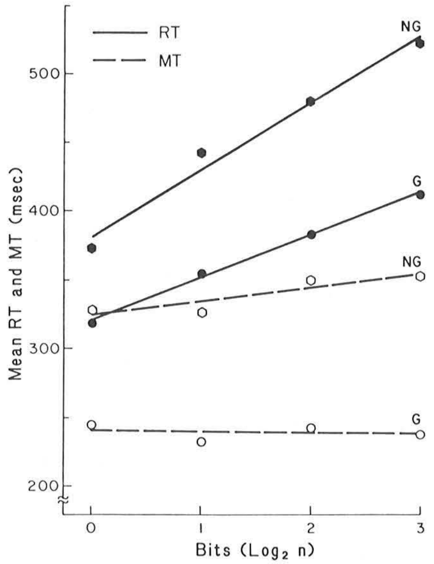

FIG. 4.7. Reaction time (RT) and movement time (MT) in NG and G groups as a function of bits of information corresponding to 1, 2, 4, and 8 light/button alternatives.

they are clearly correlated with IQ and discriminate between groups that differ in terms of generally accepted criteria of intelligence. These findings also suggest that the processes underlying *g* may be essentially simpler than their manifestations in complex problem solving and other ''real-life'' behavior, just as the cause of a disease may be simpler than its multifarious symptoms.

The speed factor that we are measuring with these tasks should not be thought of as intentional, overt speed at the level of gross behavior. It is not the kind of speed that suggests hurrying and rushing through the performance of a task. Speed can be thought of in two senses: cognitive and conative. Cognitive speed is speed of information processing. Conative speed is speed due to conscious effort, minimizing rest pauses, and the like. Conative speed as it affects performance on psychometric tests cannot begin to explain the correlation between RT and test scores. Complete abandonment of this overly simple and superficial explanation is long overdue. In our own work, we have taken pains to minimize the speed factor in test taking. All psychometric tests are given without time limit; subjects are urged to take their time and to attempt every item. We have also found that when tests were given with a time limit and scored and then subjects were given as much additional time as they felt they needed to earn a maximal score, subjects remained in approximately the same rank order under both methods of scoring, so that the correlation of the scores with another variable would be scarcely affected whether the test was timed or untimed. Also, we have found that speeded tests show no higher correlations with RT tasks than untimed tests. Clerical checking tests, which are the most dependent on speed, have the lowest *g* loadings and the poorest correlations with RT measures. For example, the Coding test, the most speed-dependent test of the ten tests in the ASVAB battery, has the lowest *g* loading in this battery and the lowest correlation with the general speed factor extracted from a battery of eight RT tests (Vernon & Jensen, 1984). The same thing is true of the speeded Coding (or Digit Symbol) subtest of the WAIS (Vernon, 1983). The clincher is that we have found a correlation close to zero between individual differences in total test-taking time (under untimed conditions) and total scores on the test.

How then can we explain the correlation between RTs in ECTs and psychometric *g?*

Several well-established concepts and principles of cognitive psychology provide a rationale for the importance of a time element in mental efficiency. The first such concept is that the conscious brain acts as a one-channel or *limited-capacity* information-processing system. It can deal simultaneously with only a very limited amount of information; the limited capacity also restricts the number of operations that can be performed simultaneously on the information that enters the system from external stimuli or from retrieval of information stored in short-term or long-term memory (STM or LTM). Speediness of mental operations is advantageous in that more operations per unit of time can be executed without overloading the system. Second, there is *rapid decay* of stimulus traces and information, so that there is an advantage to speediness of any operations that must be performed on the information while it is still available. Third, to compensate for limited capacity and rapid decay of incoming information, the individual resorts to *rehearsal and storage* of the information into intermediate or long-term memory, which has relatively unlimited capacity. But the process of storing information in LTM itself takes time and therefore uses up channel space, so there is a ''trade-off'' between the storage and the processing of incoming information. The more complex the information and the operations required on it, the more time that is necessary, and consequently the greater the advantage of speediness in all the elemental processes involved. Loss of information due to overload interference and decay of traces that were inadequately encoded or rehearsed for storage or retrieval from LTM results in ''breakdown'' and failure to grasp all the essential relationships among the elements of a complex problem needed for its solution. Speediness of information processing should therefore be increasingly related to success in dealing with cognitive tasks to the extent that their information load strains the individual's limited channel capacity. The most discriminating test items would thus be those that ''threaten'' the information-processing system at the threshold of ''breakdown.'' In a series of items of graded complexity, this ''breakdown'' would occur at different points for various individuals. If individual differences in the speed of the elemental components of information processing could be measured in tasks that are so simple as to rule out ''breakdown'' failure, as in the various RT paradigms we have used, it should be possible to predict individual differences in the point of ''breakdown'' for more complex tasks. This is the likely basis for the observed correlations between RT variables measured in relatively simple tasks and total scores on complex *g*-loaded tests.

The speed of elemental information processing may not be the most basic source of individual differences in intelligence but may be only a secondary phenomenon, derived from a still more basic source of individual differences—a hypothetical construct I have termed "neural oscillation," which would account for individual differences in intertrial variation in RT as well as in individual differences in RT averaged over a given number of trials (Jensen, 1982a, pp. 6–10). Eysenck (1982a) also regards differences in mental speed and RT as derivative, in the sense that a person's average RT is not directly attributable to the speed of neural conduction or synaptic transmission. He hypothesizes that speed differences arise from individual differences in the rate at which errors occur in the transmission of neural impulses in the cortex. The stimulus message must persist until the "pulse train" of neural impulses exceeds a certain fidelity threshold. The more random "noise" or error tendency in the neural system, the more time this takes, and hence speed of reaction is a derivative phenomenon.

So far, there has been no way empirically to decide between the hypotheses of processing speed and errors, or "noise," in the neural transmission of errors as basic to *g*. Whether these concepts will be able to account for all or only some fraction of the true-score variance in the *g* derived from a large and diverse sample of psychometric tests has yet to be determined. It will be necessary, first of all, to determine how large a correlation with *g* can be obtained from a battery of various simple chronometric tasks of sufficient number and diversity to minimize the proportion of task-specific variance in the composite score. The best composite correlations we have obtained thus far would account for at most only about half of the variance in *g*.

Multiprocess Theories

Thomson's Sampling Theory of g. E. L. Thorndike (1927) was the first systematic proponent of the theory that *g* is explainable in terms of the hypothesis that human abilities consist of independent multiple bonds or neural connections acquired through experience, and that successful performance on various tests enlists somewhat different but overlapping "samples" of all the myriad bonds that constitute ability. Thorndike believed that individuals differ innately in the potential number of bonds they can acquire, the total number being limited by the number and degree of branching of the neural elements. As this theory proposes no inherent structure or organization of the bonds themselves, Spearman (1927, Ch. V) termed all theories of this type "anarchic."

Sir Godfrey Thomson, who spent a year's postdoctoral followship working with Thorndike, developed Thorndike's bond-sampling theory further, formalizing it mathematically in his now famous book *The Factorial Analysis of Human Ability* (1948, Ch. XX). Essentially, he showed that the correlation between two tests, X and Y, could be represented as $r_{xy} = (p_x p_y)^{1/2}$, where p is the proportion of the total pool of elements or "bonds of the mind" "sampled" by a given test. From this formulation, Thomson was able to demonstrate mathematically how both *g* and specificity could come out of the factor analysis of a number of tests that call upon different but overlapping samples of elements. Thomson's sampling theory, as it has come to be known, is illustrated in Fig. 4.8. It can be seen that in this model the factors yielded by factor analysis do not represent anything in the mind, which consists only of innumerable disparate bonds or elements of some kind. The organization or structure represented by factors is seen as an artifact of the tests, which can be devised to sample large or small numbers of elements. Complex tests would sample more elements than simple tests, and complex tests would therefore be apt to be more highly correlated with other tests, and consequently would be more *g* loaded. To simulate the typical results of Spearman's factor analyses, the sampling model only requires, in Thomson's (1948) words,

> that it be possible to take our tests with equal ease from any part of the causal background; that there be no linkages among the bonds which will disturb the random frequency of the various possible combinations; in other words, that there be no "faculties" in the mind. . . . The sampling theory assumes that each ability is composed of *some but not all* of the bonds, and that abilities can differ very markedly in their "richness," some needing very many "bonds," some only a few. (p. 324).

Thomson left the number and nature of the hypothetical bonds, or elements, of the sampling theory completely unspecified. This deficiency is the core of the theory's weakness in terms of its testability as empirical science. It can be proved mathematically that any number of composite aggregates of whatever degree of correlation with each other can always be expressed as functions of elements that are themselves uncorrelated (Spearman, 1927, p. 59). Despite its superficial plausibility, Thomson's sampling theory does not qualify as a scientific theory. Although it has enjoyed much greater uncritical popularity in recent years than

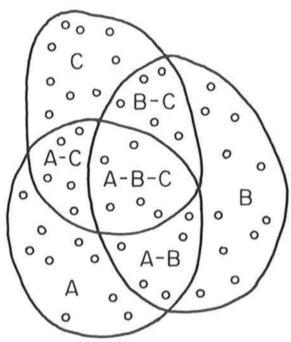

FIG. 4.8. Illustration of Thomson's sampling theory of abilities, in which the small circles represent elements or bonds and the large circles represent tests that sample different sets of elements (labeled A, B, and C). Correlation between tests is due to the number of elements sampled in common, represented by the areas of overlap.

Spearman's theory of "mental energy," it has been no more fruitful in advancing empirical research on the nature of *g* or intelligence. Loevinger's (1951) verdict seems inescapable:

> The sampling theory hardly qualifies as a true theory, for it does not make any assertion to which evidence is relevant. Perhaps the large number of adherents to this view is due to the fact that no one has offered evidence against it. But until the view is defined more sharply, one cannot even conceive of the possibility of contrary evidence, nor, for that matter, confirmatory evidence. A statement about the human mind which can be neither supported nor refuted by any facts, known or conceivable, is certainly useless. Bridgman and other philosophers of science would probably declare the sampling theory to be meaningless. (pp. 594–95)

Along with Spearman's theory of "mental energy," Thomson's rival sampling theory can be consigned to the museum of psychology's past history, but unlike phlogiston, without ever having enjoyed the scientific virtue of being empirically disproved.

Modern descendants of the sampling theory are scarcely more definite as to the number and nature of the sampled elements. A number of modern theorists conceive of intelligence, or *g*, as the entire repertoire of an individual's knowledge, skills, and problem-solving strategies available at a given point in time (e.g., Humphreys, 1984; Tyler, 1976, pp. 24–25; Undheim, 1981c). In the same key, the *g* factor has also been attributed to individual differences in the number of well-learned cognitive skills that generalize across a broad spectrum of problem-solving situations.

All theories of this type run into difficulty with the empirical finding that a relatively small variety of tests, which can in no way be construed as a representative sample of the entire repertoire of knowledge, skills, and strategies, are capable of measuring *g*. One obviously does not require a sample of the entire repertoire of knowledge, skills, and strategies to measure *g*. A few relatively content-free tests of the "fluid *g*" variety are even more *g* loaded than are tests that aim to sample individuals' entire cognitive repertoire. It is also hard to see how these theories can accommodate the substantial correlations between RT measures derived from quite simple ECTs and psychometric *g*. What repertoire is sampled by these ECTs, most of which seem entirely too elementary to be described in terms of "knowledge, skills, and strategies"? If most of the *g* variance could be predicted by chronometric measures on a number of ECTs, or by a physiological measure such as the evoked potential, which involves no conscious behavioral aspects at all, these neo-Thomsonian sampling theories (perhaps better termed "repertoire" theories) would be empirically falsified in terms of any of their meaningful implications.

Component Process Theories of g. Process theories of *g* are essentially sampling theories, but with an important difference from Thomson's bond-sampling theory and from theories that identify *g* with the entire repertoire of knowledge, skills, and strategies. The essential difference is that process theories posit some limited number of basic information-processing components, each of which can be described in terms of the particular functions it performs—functions that, when viewed in isolation, are usually too elemental to be thought of as skills or strategies at the level of overt behavior. An *information-processing component* is itself a hypothetical construct, defined as a process that operates on sensory inputs or internal representations of objects or symbols. These elementary cognitive processes have been described by terms such as stimulus apprehension, sensory encoding, iconic memory, short-term memory (STM), memory scanning, retrieval of information from long-term memory (LTM), transformation of encoded information, transfer, discrimination, generalization, eduction and mapping of relations, visualization and mental rotation of figures in 2- or 3-dimensional space, and response execution. A less elemental class of operations are *metaprocesses,* which are acquired strategies for selecting, combining and using the elementary processes, problem recognition, rule application, planning, organization of information, time allocation, and monitoring of one's own performance.

Processing theory explains psychometric g in terms of a small number of components or metacomponents that are required for performance in an extremely broad variety of tests. Individual differences in the presence or absence or efficiency of operation of these general or common components and metacomponents are what account for the positive intercorrelations among practically all complex mental tests and the consequent emergence of g when all the intercorrelations are factor analyzed. The interpretation of g in terms of componential theory has been quite thoroughly explicated by Sternberg and Gardner (1982).

Figure 4.9 depicts the hypothesized relationship between the processing variables and psychometric variables. The horizontal dashed line in Fig. 4.9 separates the behaviorally measurable or inferred psychological variables (above the line) from those that are measurable only physiologically, such as evoked brain potentials, or inferred physiological processes, such as cortical conductivity (Klein & Krech, 1952), synaptic errors (Hendrickson, 1982), neural oscillation (Jensen, 1982a), and the like. The physiological level is represented as one general factor, g_B (B for "biological"), although, given our present state of knowledge, this level could just as well be represented as several distinct physiological processes or as correlated processes, due to their sharing one common process, i.e., g_B. The nature of this physiologic underpinning of human abilities is a major focus of Eysenck's (1982b) theorizing about the findings of correlations between features of the average evoked potential and psychometric g, or g_P, which is depicted in the hexagon at the top of the hierarchy in Fig. 4.9. All of the solid lines in the figure represent correlations. (Correlations could also be shown between elements at every level and every other level of the hierarchy, but these have been omitted for the sake of graphic simplicity.)

The various elementary cognitive processes (P) are correlated through their sharing of common physiological processes. Different parts of the brain or differ-

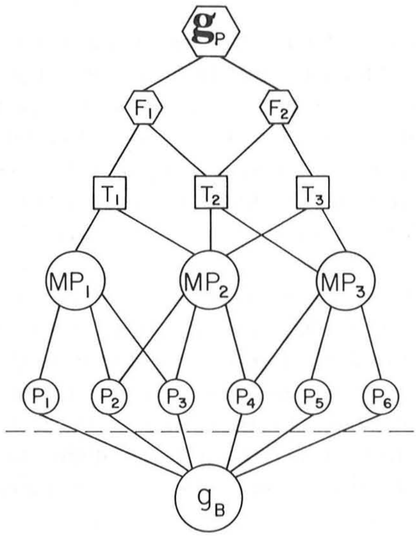

FIG. 4.9. Simplified representation of hypothesized relationships among processing components and psychometric variables: g_P—psychometric g, F—group factor, T—test, MP—metaprocesses, P—elementary cognitive process, g_B—a general biologic substrate, as reflected in physiological indices such as evoked potentials.

ent neural assemblies are presumably specialized for various aspects of information processing. The processes in this model, depicted here as being closely connected with some biological substrate, can all be measured by means of chronometric tasks, either directly or through derived scores. By subtraction of response latencies of simple tasks from the latencies of more complex tasks, one can measure individual differences in the additional processes involved in the latter.

Different sets of elementary processes, *P*, can be utilized by a given *metaprocess* (*MP*). Because metaprocesses are further removed from the biologic substrate and are probably mainly products of learning and practice, their connection to the biologic substrate is via the elementary processes which enter into the metaprocesses. Different metaprocesses are intercorrelated because they share certain elementary processes in common and also because the experiential factors which inculcate metaprocesses are correlated in the educational and cultural environment. It is probably at the level of metaprocesses that cultural differences have their primary impact.

Both processes and metaprocesses enter into performance on complex *psychometric tests* (*T*). Even a single complex test item may well depend on a number of *P*s and *MP*s for successful performance. Various tests are intercorrelated, moreover, not only because they share certain common *P*s and *MP*s, but also because they may share common information stored in long-term memory. Note that at each level in this hierarchy, something new is added in terms of environmental inputs. The cumulative impact of these acquired elements is at its maximum at the level of single items in psychometric tests. Item variance is largely *specificity*, which may arise from individuals' idiosyncratic experiences, making for unique and uncorrelated bits of information, or from complex and unique interactions among the P and MP demands and the informational content of a particular test item. In fact, all primary psychological measurements are saturated with task-specific variance. Chronometric measurements of elementary processes in specially contrived laboratory tasks are no exception. Specificity, which is the bane of individual differences research, can be reduced only by using composite scores or factor scores (which are a particular weighted composite of the component scores) derived from a number of varied tasks or tests, thereby "averaging out" the specificity of the individual tasks.

The top part of the hierarchy in Fig. 4.9, including T, F, and g_P, encompasses the realm of traditional psychometrics, including various test scores and hierarchical factors extracted by factor analysis. Here, for the sake of simplicity, are represented only two first-order factors (F_1 and F_2) and one second-order factor, psychometric g, or g_P. (The most general factor, of course, may emerge as a third-order or other higher-order factor.) Each successively higher factor level excludes some sources of variance. The primary factors, for example, exclude the test-specific variance, and the second-order factors exclude the variance that is peculiar to each primary factor, and so on. The most general factor, g_P, is the variance common to all the sources below it in the hierarchy.

Some homogeneous tests, such as Raven's Progressive Matrices, contain relatively little specificity and are therefore quite good measures of g_P. Other tests, like the Wechsler scales, although containing quite heterogeneous items and subtests with considerable specificity, yield composite scores from which, in effect, the specificity is "averaged out," providing a good measure of g_P.

Superficially very different tests, such as Verbal Analogies, Digit Span, and Block Designs, are intercorrelated presumably not because of common content or correlated educational experiences, but because they have a number of elementary processes and metaprocesses in common. Because the more superficial differences between tests contribute mainly to their specificities, these differences are not reflected in g_P. Hence it has been found that g factor scores are more highly correlated with chronometric measures of elementary processes than are any particular types of tests. Thus, although g_P and P_1, P_2, etc., appear widely separated in the schematic hierarchy, they actually seem to have greater variance overlap, as shown by the correlation, than do some of the more proximal variables. This picture may also help to elucidate the otherwise surprising finding that, although g_P is derived from factor analysis of psychometric tests which bear virtually no superficial resemblance in format, content, or method of administration to the RT techniques used in elementary cognitive tasks (ECTs), g_P shows correlations with ECTs almost as large as with the psychometric tests from which g_P is derived.

One of the crucial theoretical questions, with reference to Fig. 4.9, regarding which there is presently little consensus, is whether more of the variance in psychometric g (g_p) is attributable to the processes (P) or to the metaprocesses

(*MP*). The learned information content in the psychometric tests (*T*) can already be virtually ruled out as an important source of g variance, because tests that differ markedly in their information content, such as vocabulary and matrices, are nevertheless highly saturated with one and the same g. The multiple correlation of several simple ECTs with g_P has been so substantial in some studies as to suggest that perhaps as much as 50% of the g_P variance is accounted for by individual differences in elementary cognitive processes. If task specificity were further minimized in such studies, by using at least three or four different techniques for measuring each of the elementary processes that have already been shown to yield substantial correlations, it seems likely that even more than half of the g variance would be associated with the elementary processing variables. Also, the existing studies have not taken sufficient account of the reliability of these processing measures. Proper corrections for attenuation might appreciably raise the correlations between ECTs and g_P. Split-half or other internal consistency estimates of the reliability of ECTs usually overestimate the test-retest reliability, and it is the test-retest reliability which should be used in correcting correlations for attenuation when the correlated measurements have been obtained in different test sessions, on different days, for example, or even at different times of the same day, such as before and after lunch. Some of the ECT measurements are so highly sensitive to an individual's fluctuating physiological state from morning till night and from day to day as to have quite low test-retest reliability as compared with most psychometric tests. Theoretical interest, of course, focuses on the *true-score* multiple correlation between the elementary cognitive processes and g_P. Individual differences in metaprocesses, or strategies, might even obscure task correlations with g. Hughes (1983), for example, found that a measure of learning rate is more highly correlated ($r = -.59$, $p <$ .001) with g (i.e., Raven Matrices) when all subjects are constrained by instructions to use the same strategy for learning than when they are not so instructed and can choose their own strategies ($r = +.16$, n.s.). This is just the opposite of what one should predict if metacomponents (strategies) were the chief sources of variance in learning rates or in g. One goal of componential research is to determine the proportions of variance in g accounted for by each of a number of clearly identifiable processes and metaprocesses. This has not yet been accomplished.

There is a crucial difference between *factors* and *processes* that is often overlooked. Factors arise completely out of individual differences, and factors, including g, reflect only individual differences in whatever causal mechanisms are involved in the factors. Because of their exclusive dependence on variance, therefore, factors do not necessarily represent the operating principles of the mind. Processes that were so essential to individual survival in the course of human evolution as to be left with little or no genetic variance would not show up as factors. As far as I know, it has not been determined if there are any cognitive processes of this nature, that is, processes that might show age differences but not reliable individual differences among biologically normal, healthy persons. It is at least a safe assumption that various processes may differ in the extent of their individual differences variance, and this can be assessed when individual differences are measured chronometrically, since such measures are on a ratio scale, which permits comparisons of variability based on the coefficient of variation ($V = \sigma/\mu$). The point is that *processes* will be reflected in *factors* in proportion to their coefficients of variation. (For example, the g factor is always smaller, relative to other factors, in college students than in the general population, because students are selected essentially on g.) Unlike a factor, a process can be identified and its importance in the mental economy assessed without need to take account of individual differences. RT is measured on a task (e.g., simple RT) which it is hypothesized requires processes A and B, and RT is measured on a task (e.g., choice RT) which requires processes A, B, and C. The difference in milliseconds beween the mean RTs on the two tasks is taken as evidence for process C and indicates its magnitude in relation to other processes assessed by the same type of experimental paradigm, known as the *subtraction* method, originated by Donders (1868–69/1969) in the early years of mental chronometry. The processes that best account for g will not necessarily be those that experimental cognitive research determines are the most important in terms of their mean effects, but those on which there is the largest variance. These two features of processes may or may not be related.

Although Sternberg believes that the bulk of g is attributable to variance in metaprocesses, this view is not an essential feature of componential theories in general. Moreover, its truth has not yet been demonstrated. A proper test would logically require that an adequate number of measures of elementary cognitive processes be entered first into the stepwise multiple regression, followed by the metaprocess measures, for predicting g factor scores, thereby determining the independent contribution of metaprocess to the variance in g. The outcome of such a study would be of great theoretical importance. My guess at this point is that Sternberg's belief is wrong, and that most of the g variance will be accountable in terms of elementary cognitive processes, with little if any variance left for the residualized metaprocesses. I conjecture that the opposite would be found for many narrow group factors or, in particular, certain types of tasks that lend themselves to various strategies. A lack of some clear demarcation between processes and metaprocesses would invite further debate. Studies permitting "strong inference" are most needed.

If processes (or metaprocesses) are uncorrelated, then, of course, we must explain g in terms of a number of common processes that enter into performance on a wide variety of tests. This seems to be the gist of Sternberg's componential theory of g (Sternberg & Gardner, 1982). But if the processes themselves are correlated with each other and yield a g much like psychometric g, then the theoretical picture is quite different. How do we explain the correlations between the process measures? In terms of still other, even more elemental, processes?

And what if they too are correlated? How far down the reductionist hierarchy will this "infinite regress" extend?

There is every indication that elementary cognitive processes are, in fact, quite highly correlated. This fact has frustrated some of Sternberg's componential analyses, the clarity of which depends on there not being very high correlations between measures of putatively different processes. For example, Sternberg and Gardner (1982, p. 249), using chronometric techniques, measured individual differences in three different tasks which were intended to yield parameter estimates of three distinct components. But the three tasks (analogies [A], classification [C], and series completion [S]) were all so highly correlated ($r_{AC} = .86$, $r_{AS} = .85$, $r_{CS} = .88$) that when the common factor was partialled out, the little remaining variance attributed to the residualized components was unreliable. The loadings of the three tasks on their common factor are A = .91, C = .94, S = .93, without correction for attenuation. It leaves one to wonder if there are individual differences in components independent of the common factor, which may be the ubiquitous *g*. Sternberg himself has specifically noted that when the time taken for each of the component processes in his chronometric analogies tasks are factor analyzed with psychometric reference tests of *g*, individual differences in the average time for all the components (what Sternberg calls the regression constant) show a higher correlation with *g* than any of the single component latencies. Sternberg (1979a) writes:

> Information-processing analyses of a variety of tasks have revealed that the "regression constant" is often the individual differences parameter most highly correlated with scores on general intelligence tests. This constant measures variation that is constant across all of the item or task manipulations that are analyzed via multiple regression. The regression constant seems to bear at least some parallels to the general factor. (p. 24)

Referring to the same point elsewhere, Sternberg (1979b) says this about the "regression constant": ". . . we can feel pleased to be rediscovering Spearman's *g* in information processing terms." This is not an admission of failure for the componential theory of *g*, but an important discovery for which Sternberg deserves credit. But it also suggests that the search for *g* has to be pushed below the level of metaprocesses and elementary cognitive processes. Look again at where that leads us in terms of Fig. 4.9. *Any* kind of sampling theory, at least at the level of cognitive processes, may prove wholly unnecessary for explaining *g*. Do people differ in psychometric *g* because they are strong or weak on different components? Or is the *g* of the processing components essentially the same as psychometric *g?* Although there are distinctly different information processes, as demonstrated in experimental mental chronometry (e.g., Posner, 1978), individual differences in these processes may be very highly correlated because of some general property of the nervous system that acts in all of them.

One of the best known ECTs, the S. Sternberg short-term memory scanning paradigm (S. Sternberg, 1966, 1975), can be used to illustrate the problem of seeking the explanation of g in terms of tests sampling a number of elementary cognitive processes that are common to many tests, but which are themselves so saturated with some common source of variance, perhaps the same g they are intended to explain, as to force us to seek the explanation of g at a still more basic level of analysis. In the Sternberg memory-scan (M-scan) paradigm, the subject is shown (either simultaneously or sequentially) a set of digits, varying in set size (s) from 1 to 7 digits. After the subject has studied the series (termed the *positive set*) for a few seconds, the set disappears, and 1 or 2 seconds later a single target digit appears on the screen. The subject responds as quickly as possible by pressing buttons labeled either YES or NO in terms of whether the target digit was or was not a member of the positive set. The subject's RT is measured in milliseconds. Numerous studies have shown that it takes slightly longer to respond NO than YES, and RT increases as a linear function of set size. (The serial position of the target digit in the positive set has no effect on the RT.) Studies have also shown that the intercept and slope of this function, or the overall mean RT, are negatively correlated with psychometric g (e.g., Chiang & Atkinson, 1976; Keating & Bobbitt, 1978; McCauley, Dugas, Kellas, & DeVellis, 1976).

The *intercept* of the linear function relating RT to set size reflects E, the time required for encoding the target digit; B, the time for making a binary decision (Yes or No); and R, response production (releasing or pressing a button). The *slope* of the function reflects S, the speed of scanning short-term memory, specifically the time required per digit. A subject's mean RT for any given set size is hypothesized to comprise the time required for each of the information-processing components (i.e., E, B, R, S).

The reverse of this M-scan paradigm is called visual scan (V-scan). Everything is exactly the same except that the single target digit is presented *first*, followed by the positive set. The subject must visually scan the positive set and respond YES or NO as to the presence or absence of the target digit in the positive set. No scanning of STM is involved, just visual scanning of the physically displayed set of digits.

Visual scanning and STM memory scanning are obviously completely different processes. Yet in the four studies in which both the V-scan and M-scan paradigms have been used with the same group of subjects, there were no significant differences between V-scan and M-scan in intercepts, slopes, or overall mean RT (Ananda, 1985; Chiang & Atkinson, 1976; Gilford & Juola, 1976; Wade, 1984). But the really important point, in terms of implications for the componential sampling theory of g, is the finding that individual differences in the RT parameters are very highly correlated across the V-scan and M-scan tasks, so much so, in fact, as to swamp the possibility of demonstrating any independent abilities in the two types of task. Ananda (1985) found a correlation of $+.69$ between mean RTs on M-scan and V-scan; Wade (1984) found a correlation of +.85. There is no telling how much higher these correlations would be if they could be corrected for attenuation. (Neither study determined test-retest reliability.) Chiang and Atkinson (1976) gave their subjects more trials and therefore obtained considerably more reliable measurements of individual differences. Their correlation between V-scan and M-scan was +.97 for intercepts and +.83 for slopes. These very high correlations (not corrected for attenuation) were obtained despite the restricted range of ability in the Stanford University students who served as subjects. (Corrected for attenuation [using Day 2-Day 3 test-retest reliability], the above correlations are 1.20 and 1.13, respectively.) Chiang and Atkinson state, "It might be argued that performance on these search tasks is related to a general factor, speed, and that it is not useful to break down performance into several component processes or to distinguish between parameters of these processes" (p. 668). But this conclusion is a nonsequitur. Distinctly different processes may be involved in M-scan and V-scan, but the different processes may not be distinguishable in terms of individual differences because some more basic general factor that affects speed in all cognitive operations is common to both processes. In fact, we generally find such high correlations among the RTs to various ECTs that only one factor accounts for nearly all of the intercorrelation among the ECTs. Nonspeeded psychometric tests of *g* also have considerable loadings on the same general speed factor.

If the condition I have described with respect to the M-scan and V-scan tasks is found in future research to be generally typical of most other ECTs that presumably involve distinctly different processes, and if it is their largest common factor, rather than any subordinate factors, that is correlated with psychometric *g*, it would seem clear that an adequate theory of *g* will most probably have to invoke some even more basic level of analysis than is provided by the processing-component sampling theory. It seems likely that continuing effort to achieve a scientifically adequate theory of one of the most controversial psychological constructs will force it out of psychology altogether and arrive at an empirically testable formulation in genuinely physiological terms. But this may be the ultimate fate of any truly important construct of psychology. Is it not the ultimate "psychologists' fallacy" to be satisfied with a psychological explanation of a psychological phenomenon?

REFERENCES

Agrawal, N., Sinha, S. N., & Jensen, A. R. (1984). Effects of inbreeding on Raven Matrices. *Behavior Genetics, 14*, 579–585.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*, 1289–1292.

Ananda, S. M. (1985). *Speed of information processing and psychometric abilities in later adulthood.* Unpublished doctoral dissertation, University of California, Berkeley.

Block, J. B. (1968). Hereditary components in the performance of twins on the WAIS. In S. G. Vandenberg (Ed.), *Progress in human behavior genetics*. Baltimore, MD: Johns Hopkins.

Burt, C. (1940). *The factors of the mind: An introduction to factor analysis in psychology*. New York: Macmillan.

Burt, C. (1961). Factor analysis and its neurological basis. *British Journal of Statistical Psychology, 14*, 53–71.

Carlson, J. S., & Jensen, C. M. (1982). Reaction time, movement time, and intelligence: A replication and extension. *Intelligence, 6*, 265–274.

Carlson, J. S., Jensen, C. M., & Widaman, K. F. (1983). Reaction time, intelligence, and attention. *Intelligence, 7*, 329–344.

Carroll, J. B. (1980). *Individual difference relations in psychometric and experimental cognitive tasks*. Chapel Hill, NC: L. L. Thurstone Psychometric Laboratory, University of North Carolina.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

Chiang, A., & Atkinson, R. C. (1976). Individual differences and interrelationships among a select set of cognitive skills. *Memory and Cognition, 4*, 661–672.

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.

Donders, F. C. (1969). Over de snelheid van psychische processen. In W. G. Koster (Ed. and Trans.), *Attention and performance II, Acta Psychologia, 30*, 412–431. (Original work published 1868–69)

Eysenck, H. J. (1939). Primary mental abilities. *British Journal of Educational Psychology, 9*, 270–275.

Eysenck, H. J. (Ed.). (1982a). *A model for intelligence*. Heidelberg: Springer-Verlag.

Eysenck, H. J. (1982b). The psychophysiology of intelligence. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & V. Willson (Eds.), *Methodological and statistical advances in the study of individual differences*. New York: Plenum.

Fogarty, G. J. (1984). *The structure of abilities underlying performance on competing tasks*. Unpublished doctoral dissertation, University of Sydney.

Fogarty, G. J., & Stankov, L. (1982). Competing tasks as an index of intelligence. *Personality and Individual Differences, 3*, 407–422.

Garrett, H. E., Bryan, A. I., & Perl, R. E. (1935). The age factor in mental organization. *Archives of Psychology* (No. 176).

Gilford, R., & Juola, J. (1976). Familiarity effects on memory search and visual search. *Bulletin of the Psychonomic Society, 7*, 142–144.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.

Haier, R. J., Robinson, D. L., Braden, W., & Williams, D. (1983). Electrical potentials of the cerebral cortex and psychometric intelligence. *Personality and Individual Differences, 4*, 591–599.

Hemmelgarn, T. E., & Kehle, T. J. (1984). The relationship between reaction time and intelligence in children. *School Psychology International, 5*, 77–84.

Hendrickson, A. E. (1982). The biological basis of intelligence. Part I: Theory. In H. J. Eysenck (Ed.), *A model for intelligence*. Heidelberg: Springer-Verlag.

Hughes, O. L. (1983). A comparison of error based and time based learning measures as predictors of general intelligence. *Intelligence, 7*, 9–26.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3,* 105–120.

Humphreys, L. G. (1984). General intelligence. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on Bias in Mental Testing* (pp. 221–247). New York: Plenum.

Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal, 17,* 153–170.

Jensen, A. R. (1982a). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). Heidelberg: Springer-Verlag.

Jensen, A. R. (1982b). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jensen, A. R. (1983). Effects of inbreeding on mental-ability factors. *Personality and Individual Differences, 4,* 71–87.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures, 7,* 93–118.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193–219.

Jensen, A. R., & Munro, E. (1979). Reaction time, movement time and intelligence. *Intelligence, 3,* 121–126.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability differences on the WISC-R. *Personality and Individual Differences, 3,* 423–438.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187–200.

Keating, D. P., & Bobbitt, B. (1978). Individual and developmental differences in cognitive processing components of mental ability. *Child Development, 49,* 155–169.

Klein, G. S., & Krech, D. (1952). Cortical conductivity in the brain-injured. *Journal of Personality, 21,* 118–148.

Lansman, M., Poltrock, S. E., & Hunt, E. (1983). Individual differences in the ability to focus and divide attention. *Intelligence, 7,* 299–312.

Loevinger, J. (1951). Intelligence. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand.

McCauley, C., Dugas, J., Kellas, G., & DeVellis, R. F. (1976). Effects of serial rehearsal training on memory search. *Journal of Educational Psychology, 68,* 474–481.

Macphail, E. M. (1985). Comparative studies of animal intelligence: Is Spearman's *g* really Hull's *D? Behavioral and Brain Sciences, 8,* 234–235.

Maxwell, A. E. (1972). Factor analysis: Thomson's sampling theory recalled. *British Journal of Mathematical and Statistical Psychology, 25,* 1–21.

Nagoshi, C. T., & Johnson, R. C. (1986). The ubiquity of *g*. *Personality and Individual Differences, 7,* 201–207.

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

Posner, M. I. (1978). *Chronometric exploration of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Price, B. (1936). Homogamy and the intercorrelation of capacity traits. *Annals of Eugenics, 7,* 22–27.

Schafer, E. W. P. (1985). Neural adaptability: A biological determinant of *g* factor intelligence. *Behavioral and Brain Sciences, 8,* 240–241.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.

Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.

Spearman, C. E. (1973). *The nature of 'intelligence' and the principles of cognition.* New York: Arno Press. (Original work published 1923)

Stankov, L. (1983). The role of competition in human abilities revealed through auditory tests. *Multivariate Behavioral Research Monographs* (No. 83-1).

Sternberg, R. J. (1979a). *Components of human intelligence* (Tech. Rep. No. 19). Office of Naval Research.

Sternberg, R. J. (1979b). In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement.* Norwood, NJ: Ablex.

Sternberg, R. J., & Gardner, M. K. (1982). A componential interpretation of the general factor in human intelligence. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 231–254). Heidelberg: Springer-Verlag.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153,* 652–654.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32.

Tambs, K., Sundet, J. M., & Magnus, P. (1984). Heritability analysis of the WAIS subtests. A study of twins. *Intelligence, 8,* 283–293.

Thomson, G. H. (1948). *The factorial analysis of human ability* (3rd ed.). Boston: Houghton-Mifflin.

Thorndike, E. L. (1927). *The measurement of intelligence.* New York: Teachers College, Columbia University.

Thurstone, L. L. (1935). *Vectors of mind.* Chicago: University of Chicago Press.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs* (No. 1). Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tyler, L. E. (1976). The intelligence we test—an evolving concept. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 13–26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Undheim, J. O. (1981a). On intelligence I: Broad ability factors in 15-year-old children and Cattell's theory of fluid and crystallized intelligence. *Scandanavian Journal of Psychology, 22,* 171–179.

Undheim, J. O. (1981b). On intelligence II: A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. *Scandanavian Journal of Psychology, 22,* 181–187.

Undheim, J. O. (1981c). On intelligence IV: Toward a restoration of general intelligence. *Scandanavian Journal of Psychology, 22,* 251–265.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7,* 53–70.

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5,* 411–423.

Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.

Wade, J. P. (1984). *The relationships among intelligence and six chronometric paradigms with academically able students.* Unpublished doctoral dissertation, Arizona State University.

APPENDIX

Two types of RT apparatus were used. The first is shown in Figure A. Templates are placed over the console, exposing either 1, 2, 4, or 8 of the light-button combinations. When one of the lights goes on, the subject removes his finger from the central home button and presses a button adjacent to the light, which puts out the light. Fifteen trials are given at each level of complexity—1, 2, 4, or

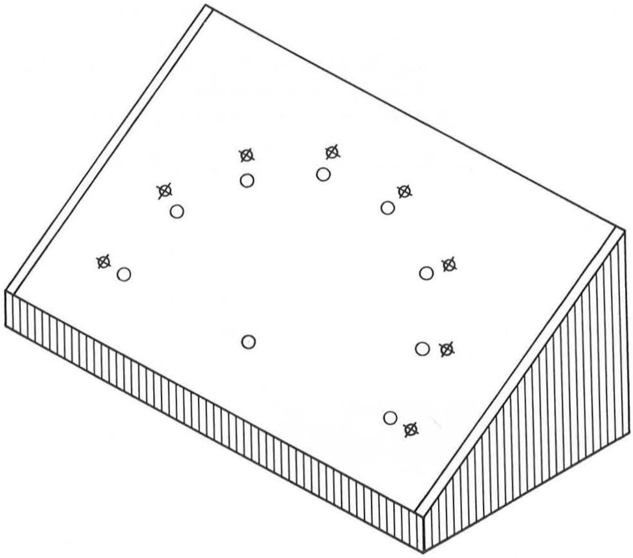

FIG. A. Subject's console of the reaction time aparatus. Pushbuttons indicated by circles, green jewelled lights by circled crosses. The "home" button is in the lower center.

8 light-buttons. RT is the time taken to get off the home button after one of the lights goes on. I shall refer to this task simply as the *RT task* (RT). The other tasks all use a two-choice console pictured in Figure B. In the *Memory Scan* task (DIGIT), a set of digits consisting of anywhere from 1 to 7 digits is simultaneously presented for 2 seconds on the display screen. After a 1-second interval, a single probe digit appears on the screen. The subject's task is to respond as quickly as possible, indicating whether or not the probe was a member of the set that had previously appeared by raising his index finger from the home button and pushing one of the two choice buttons labeled "yes" and "no." The subject's RT is the interval between the onset of the probe digit and the subject's releasing the home button. The subject's score (the average of his RTs to 84 such digit sets) provides a measure of the speed of short-term memory processing, that is, the speed with which information held in short-term memory can be scanned and retrieved.

The *Same–Different* tasks (SD2) measures the speed of visual discrimination of pairs of simple words that are *physically* the same or different, for example, DOG–DOG or DOG–LOG. The instant that each of 26 pairs of the same or different words is presented, the subject raises his finger from the home button and presses one of the two choice buttons labeled *S* (same) and *D* (different).

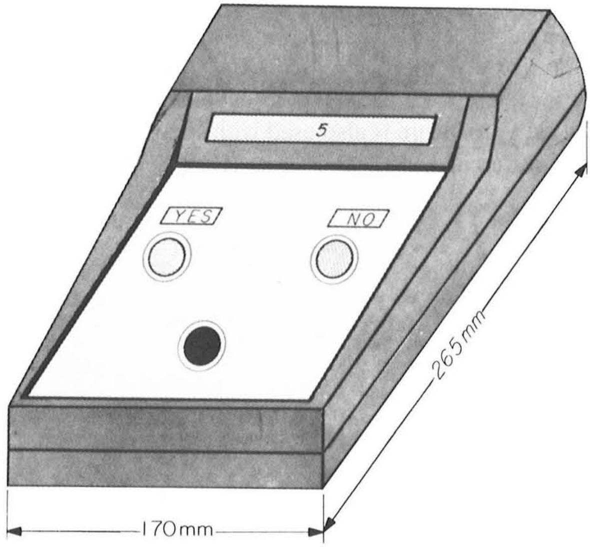

FIG. B. Subject's console used for the digit memory scan, physically same-different words, and synonyms-antonyms test, showing display screen, the two-choice response buttons, and the "home" button (lower center).

Again, the subject's RT is the average interval between onset of the word pair and releasing the home button.

The *Synonym–Antonym* task (SA2) works much the same way, but in this test pairs of words are presented that are semantically either similar or opposite in meaning, for example, BIG–LARGE or BIG–LITTLE. All the synonyms and antonyms are composed of extremely common, high-frequency words, and all items can be answered correctly by virtually any third-grader under nonspeeded test conditions. The only reliable source of individual differences is the speed with which the decisions are made. This task measures the subject's speed of access to highly overlearned verbal codes stored in long-term memory.

In the *Dual Processing* tasks, the subject is required to do two things, thus creating some degree of cognitive trade-off, or processing efficiency loss, between storage of information in short-term memory and retrieval of semantic information from long-term memory. In this task, we sequentially combine the digit Memory Scan task and the Same–Different task, or the Memory Scan task and the Synonyms–Antonyms task. First, the subject is presented with a set of 1 to 7 digits for 2 seconds. This presentation is immediately followed by a Same–Different (or Synonym–Antonym) word pair, and the subject must respond "same" or different" (pressing buttons labeled *S* or *D*). Next, the probe digit appears, and he must respond "yes" or "no" to indicate whether or not the probe was a member of the digit set shown previously. The RT (release of home button) is measured for the Same–Different responses to the words (DT2 WORDS) and for the yes–no responses to the probe digits (DT2 DIGITS). The very same dual task procedure is also used with synonyms-antonyms (in place of physically same-different words) and digits (DT3 WORDS and DT3 DIGITS).