Human Intelligence: An Introduction to Advances in Theory and Research

David F. Lohman
The University of Iowa

Recent advances in three research traditions are summarized: trait theories of intelligence, information-processing theories of intelligence, and general theories of thinking. The discussion of trait theories of intelligence focuses on the theory of fluid and crystallized abilities, particularly recent elaborations of this theory proposed by Horn (1985) and by Snow (1981). Their work provides a convenient framework for the discussion of information-processing theories of intelligence. I summarize attempts to build process theories of the major ability factors identified in Horn's (1985) version of this ability model: mental speed, verbal-crystallized abilities, fluid-reasoning abilities, and spatial-visualization abilities. I discuss Sternberg's (1985) recent attempts to develop a comprehensive theory of intelligence and ask how a theory of intelligence might be derived from the sort of general theories of thinking currently advanced in cognitive psychology and artificial intelligence (AI). The paper concludes with some speculations about the meaning of the construct intelligence and some suggestions for research on it.

What is intelligence? How does it develop? Does it decline? Has cognitive science really changed our understanding of this construct? Old questions about intelligence have been raised with a renewed vigor, and new questions have been posed. In short, there has been a remarkable resurgence of research on human abilities in the past 15 years, fueled in part by legal challenges to intelligence tests, but in even larger measure by a renewed interest in cognition in psychology. New methods of investigation and theories of cognition have been applied to old tests and theories of individual differences. Although the results have not met the loftier expectations of some advocates, progress has been made. The purpose of this paper is to provide a sampling of this progress, to note some of the problems that have attended it, and to suggest some research strategies for future research on human intelligence.

I focus on three research traditions: trait theories of intelligence, information-processing theories of intelligence, and general theories of thinking. The discussion of trait theories of intelligence focuses on Cattell's (1963) theory of fluid and crystallized abilities, particularly the elaborations of this theory proposed by Horn (1985) and by Snow (1981). Their work provides a convenient framework for the discussion of information-processing theories of intelligence. First, I summarize

---

I thank Steve Dunbar, Len Feldt, Dave Frisbee, Tom Rocklin, Bob Thorndike, Don Yarbrough, and Bill Zwick for their helpful suggestions on an earlier version of this paper, portions of which were published in R. M. Thorndike and Lohman (1989), and Bob Sternberg, David Perkins, and two anonymous reviewers for their suggestions on a more recent draft. Requests for reprints should be sent to the author at 366 Lindquist Center, The University of Iowa, Iowa City, IA 52242.
attempts to build process theories of the major factors identified in Horn’s (1985) model, such as the work of Jensen (1982) and Eysenck (1982) on mental speed, of Hunt (1985) and Frederiksen (1982) on verbal-crystallized abilities, of Sternberg (1977) on fluid-reasoning abilities, and of Pellegrino and Kail (1982) and Lohman (1988) on spatial-visualization ability. This section concludes with a discussion of Sternberg’s (1984, 1985) recent attempts to develop a comprehensive theory of intelligence. I then turn the problem around. Instead of asking how cognitive science might help us understand existing tests or ability constructs, I ask how a theory of intelligence might be derived from the sort of general theories of thinking currently advanced in cognitive psychology and artificial intelligence (AI). Here the discussion emphasizes Anderson’s (1983) ACT* theory (the latest version of his Adaptive Control of Thought system) and the “New Connectionism” of Rumelhart, McClelland, and the PDP (Parallel Distributed Processing) Research Group (1986).

The paper concludes with some speculations about the meaning of the construct intelligence and some suggestions for research on it.

*The resurgence of general ability.* Several developments converged in the early 1970s to renew interest in the construct intelligence. First, there was the growing realization that the ability profiles provided by multiple-aptitude batteries were not as useful for prediction as many had hoped (McNemar, 1964). Although there were exceptions, the predictive validities of the several scores from multiple-aptitude batteries were repeatedly found to be little better than the corresponding validity of one general factor estimated from the same battery.¹ Nor were the specific abilities that Thurstone (1938) and Guilford (1959) had identified of much use in attempts to adapt instructional methods to the ability profile of the learner. Instead, general ability accounted for most of the findings. In their summary of 20 years of research on Aptitude X Treatment interactions, Cronbach and Snow (1977) concluded:

It has become fashionable to decry the use of measures of general ability, and sometimes their use has been prohibited in school systems. The attackers usually insist that the tests do not assess ability to learn, and it is often proposed to substitute measures of achievement or “learning styles.” . . . While we see merit in a hierarchical conception of abilities, with abilities differentiated at coarse and fine levels, we have not found Guilford’s subdivision a powerful hypothesis . . . . Instead of finding general abilities irrelevant to school learning, we find nearly ubiquitous evidence that general measures predict amount learned or rate of learning or both. And, whereas we had expected specialized abilities rather than general abilities to account for interactions, the abilities that most frequently enter into interactions are general. Even in those programs of research that started with specialized ability measures and found interactions with treatment, the data seem to warrant attributing most effects to general ability. (pp. 496–497)

Thus, on one hand, special abilities failed either to predict educational outcomes better than general ability or to predict which students would profit from specialized educational interventions designed to match their particular patterns of abilities. On the other hand, American theorists gradually adopted a hierarchical model of abilities which, while allowing for both broad and narrow abilities, clearly emphasized the role of general ability.

*The cognitive revolution.* The second development was an outgrowth of the cognitive revolution in psychology. From Watson (1925) until Skinner (1953),
American psychology was dominated by the belief that mind was not the proper subject matter for psychology. Studies of animal learning or conditioning were the norm. Thinking and reasoning were considered complex behaviors that would be explained sometime in the future after elementary mechanisms of learning were adequately understood. By the mid-1960s, however, this promise was wearing thin. Psychology seemed not to be building toward the explanation of complex phenomena but, if anything, was digging increasingly deeper into reductionism. Some had already called for a rejection of radical behaviorism on theoretical grounds (Chomsky, 1959). But it was the emergence of the computer as a metaphor for mind and as a vehicle for testing theories about thinking that finally dethroned behaviorism. Rather swiftly, the mainstream of psychology moved from conditioning to perception and then to thinking and problem solving. By 1985, in the first paragraph of his introductory text on cognitive psychology, Anderson was proclaiming, "the goal of cognitive psychology is to understand human intelligence and how it works" (p. 1). Thus, in 2 decades, the word intelligence moved from the periphery of American psychology to its center.

The cognitive revolution had two rather different influences on theories of human intelligence. There were some who saw that the methods and theories of the cognitive psychologists provided a new way to understand what intelligence and other ability tests were really measuring. Carroll (1976), Glaser (1972), Hunt (e.g., Hunt, Frost, & Lunneborg, 1973), Sternberg (1977), and Snow (e.g., Snow, Marshall, & Lohman, 1976) were leaders in this effort. There were others, however, who were not at all concerned with intelligence as an individual difference construct. These investigators sought to develop theories of human cognition and, at times, to simulate their theories in computer programs that then displayed AI. Both of these efforts will be briefly reviewed in this paper.

*The Challenge of Process*

Although most research on intelligence has focused on the products of intelligence, both theoreticians and clinicians have long called for greater attention to the process of intelligent thinking.

Nobody has ever made an inventory of tasks [that define the universe of intellectual tasks], determined the correlation of each with intellect, selected an adequate battery of them, and found the proper weights to attach to each . . . If anybody did this wisely, a large fraction of his labor would be precisely to find out what abilities our present instruments did measure, and how these abilities were related to intellect; or to find out what abilities constituted intellect, and how these abilities were measured by our present instruments. (E. L. Thorndike, Bregman, Cobb, & Woodyard, 1926, p. 2)

Three decades later in his call for the unification of the two disciplines of scientific psychology—the correlational psychology of mental testing and the experimental psychology of learning—Cronbach (1957) argued

Sophistication in data analysis has not been matched by sophistication in theory. The correlational psychologist was led into temptation by his own success, losing himself first in practical prediction, then in a narcissistic program of studying his tests as an end in themselves. A naive operationism enthroned theory of test performance in the place of theory of mental processes. (p. 675)
In this Cronbach echoed Thurstone (1947), who considered a factor-analytic study of abilities only the first step in a research program. Ability factors identified in such studies should be investigated in experiments designed to manipulate and thus identify “the processes which underlie” the factors (p. 55). But such experiments had little appeal in a psychology dominated by behaviorism, and so the research program Thurstone advocated had to await the rediscovery of mental process by the mainstream of American experimental psychology.

Cognitive Science and the Computer Metaphor

Recent research on intelligence has been driven by a renewed interest in cognition in psychology and in many other fields. Cognitive science is the term now commonly used to refer to this new blend of computer science, cognitive psychology, linguistics, neuropsychology, philosophy, and instructional psychology. Although roots of the cognitive revolution may be traced to many earlier sources, several observers see 1956 as the pivotal year in the development of cognitive science. In that year, Newell and Simon (see Newell, Shaw, & Simon, 1957) reported their success in devising a computer program that could actually prove theorems in logic. In the same year, Bruner, Goodnow, and Austin published their Study of Thinking, and Miller published a seminal paper on short-term memory in which he argued that the capacity of this memory store seemed to be limited by “the magic number seven” (Newell & Simon, 1972, p. 4). The cognitive revolution gathered momentum in the 1960s and achieved ascendancy during the 1970s (see Gardner, 1985).

The computer has contributed importantly to this revolution in at least two ways. The most obvious contribution of the computer has been as a metaphor for human cognition. This metaphor has taken several forms. At the simplest level, direct analogies have been made between the hardware of the computer and the human cognitive system. Computers have devices for encoding information from external sources (card readers, keyboards), temporarily storing it (memory buffers), transforming it (central processors), retaining it on long-term storage devices (tapes, disks), and producing output (printers, video displays). Early models of human information processing relied heavily on this analogy in positing similar structures in the human cognitive system. When used in this way, the computer is but the latest mechanical metaphor for mind in psychology (Marshall, 1977). Although more sophisticated than previous metaphors such as the wax tablet or the hydraulic pump, the computer metaphor is incomplete and even misleading. For example, some researchers have begun to question the extent to which theorizing has been artificially constrained by the serial-processing, digital computer. New research programs based on parallel processing may circumvent some of these problems, particularly for modeling perception and other nonlinguistic processes. But, as will be explained, these theories have their critics too.

Some analogies between computers and human cognition go considerably beyond comparisons of the superficial characteristics of system hardware. In particular, it is argued that similar principles govern the functioning of any system that processes information. Fodor (1981) and others who espouse this computational metaphor for thought treat the mind as a device for manipulating symbols. At this level of abstraction, differences in hardware, whether electronic or neurophysiological, are thought to be irrelevant. Whether such an assumption is tenable is a hotly debated issue in cognitive science.
However, all would agree that the contribution of the computer has far exceeded its admittedly limited value as a metaphor for the human cognitive system. The greater contribution of the computer has been as a tool for developing and testing theories of cognition or, as Anderson and Bower (1973) put it, for experimenting on the nature of the connection between stimulus and response. In this way, the computer has changed the evidentiary base to include something other than human behavior. Theories of thinking and learning can be formalized as computer programs. Programs gain a measure of plausibility if they solve problems using sequences of steps that are similar to the steps used by successful human problem solvers or if, when failing to solve problems, they make errors that mimic human errors. A constant exchange between those who study human problem solving in the psychological laboratory and those who attempt to develop computer programs that display AI serves to refine and extend both efforts.

Some would object that such comparisons between humans and computers diminish human dignity. However, cognitive science makes no pretense that computational theories completely account for human cognition. Computational models of thought are in principle no different from computational models of the weather (Miller, 1981). Yet, as Miller observes, no one fears that a tornado might destroy the computer center when the computer is used to model the behavior of tornadoes. Nor do we dismiss efforts to model the weather because such models will never produce rain. Perhaps we expect more from computational models of thought because “the brain is itself a computer in a sense in which the weather is not,” and so a “computer that models an intelligent brain is expected to be a brain” (p. 220).

**Contributions of Cognitive Research**

Cognitive science has contributed to the understanding of human intelligence in three ways. First, methods and theories of cognitive science have been applied to existing tests of intelligence, either through experimental analysis of tasks taken from intelligence and other ability tests, or through careful study of the problem-solving or other information-processing characteristics of individuals identified as more or less able by existing tests. In this way, cognitive psychology offers a new source of evidence on the construct validity of tests and the ability factors they define. Second, tests of intelligence and narrow abilities are often used to predict performance in some non-test situation (e.g., conventional schooling). Careful study of the knowledge and processing demands of these criterial performances has led to the development of new measurement strategies and suggestions for the refinement of existing measures (Frederiksen, 1984; Snow & Lohman, 1989). Third, cognitive science has sought to move beyond existing definitions of intelligence grounded in individual differences to develop general theories of thinking and learning. New measures are then developed to estimate particular processes or knowledge structures hypothesized by these theories. Patterns of individual differences on these new measures are then investigated, usually by determining relationships between new measures and scores on existing tests or experimental tasks.

The following section contains a brief review of attempts to understand intelligence through the study of existing tests or ability constructs defined by such tests. Cattell’s theory of fluid and crystallized abilities has had a major impact on these efforts, particularly the theories of Horn (1985), Snow (1981), and Sternberg (1985),

337
and so his theory and recent extensions of it are summarized first. Then, experimental research on four of the major ability constructs identified by Cattell, Horn, and other theorists is summarized. The four constructs are verbal-crystallized (Gc) ability, spatial-visualization (Gv) ability, fluid-reasoning (Gf) ability, and mental speed (Gs).

Controversies About Intelligence

Controversies about the nature of intelligence seem to repeat themselves. Two of the most important controversies relate to the question of whether the general (sometimes called g) factor that is commonly equated with intelligence should be viewed as a psychological entity, or whether it is merely a mathematical abstraction. E. L. Thorndike (see E. L. Thorndike et al., 1926) and Thomson (1920) were early advocates of the view that responses to items on intelligence tests represent a particular sample of mental bonds, and thus intelligence was better understood as a mathematical abstraction than as a psychological entity. Humphreys (1985) gives a modern statement of this view. Spearman, on the other hand, interpreted the g factor as the ability to reduce relations and correlates. Sternberg’s (1977) early work on analogical reasoning constitutes a modern version of this view. This controversy has important implications for the potential contributions of cognitive theory to a theory of intelligence. If the ubiquitous general factor is simply a mathematical dimension (Humphreys, 1985), then analyses of tasks used on intelligence tests are unlikely to isolate a particular set of mental processes that are the core of intelligence. In fact, tests that are a good measure of this dimension should be composed of maximally heterogeneous items and thus would be psychologically complex (Humphreys, 1986). However, higher order processes such as coordination of existing routines or assembly of new routines (Snow, 1981) might still emerge across diverse performances (see Butcher, 1968, p. 25).  

The second controversy, often correlated with the first, is whether intelligence is an innate cognitive capacity or, instead, an acquired set of cognitive competencies. Hereditarians such as Burt (1958), Terman (1922), Jensen (1980), and Eysenck (1982) argue that good intelligence tests are—or should be—measures of this basic, biologically-based capacity. Others, such as Humphreys (1986) and Cronbach (1972), claim that potential and capacity are pie-in-the-sky concepts with no place in a scientific account of human ability. In fact, both argue that the psychology of individual differences would be well rid of the term intelligence. This controversy is reflected in the search for neurological correlates of intelligence test scores among the hereditarians and perhaps in the search for an explanation of intelligence in terms of structural differences (e.g., capacity of working memory, rate of information transfer in memory) by like-minded cognitive psychologists. On the other hand, those who believe that abilities are acquired competencies tend to emphasize the importance of knowledge in thinking (Glaser, 1984), to study the development of abilities rather than attempt to explain individual differences at a particular point in time (Kail & Pellegrino, 1985), and to view intelligence as a product of formal schooling, not simply as a predictor of success in that medium (Snow & Yalow, 1982).

The third perennial controversy concerns the question of whether intelligence is unitary, as Spearman emphasized, or has multiple dimensions, as E. L. Thorndike, Thurstone, and Guilford emphasized.
The Theory of Fluid and Crystallized Abilities

It is fitting that the most popular current resolution to the debate between Spearman, Thorndike, and Thurstone about the dimensions of intelligence was proposed by an Englishman who received his PhD under Spearman (in 1929), completed a postdoctoral fellowship under E. L. Thorndike (in 1937), conducted research with both Burt and Thurstone (Cattell, 1971, p. ix), and eventually took up permanent residence in the United States. In 1941, shortly after accepting a position at Harvard, Cattell proposed a quasi-hierarchical model of human abilities with two general factors at the apex (rather than the one advocated by Spearman). Each was defined by several of the primary factors Thurstone had identified. Cattell called these two factors fluid intelligence (Gf) and crystallized intelligence (Gc).

In the earliest published account of the theory, Cattell (1943) argued that fluid ability was “a purely general ability to discriminate and perceive relations between any fundamentals, new or old” (p. 178). Fluid ability was hypothesized to increase until adolescence and then slowly decline. It was thought to represent the “action of the whole cortex” (p. 178). Further, fluid intelligence was thought to be the cause of the general factor found among ability tests administered to children and among the “speeded or adaptation-requiring” (p. 178) tests administered to adults. Crystallized intelligence, on the other hand, was thought to consist of “discriminatory habits long established in a particular field” that were originally acquired through the operation of fluid ability but that no longer required “insightful perception” (p. 178). The empirical facts Cattell hoped to explain by this theory were the relative independence of individual differences in speed and power in adult intellectual performance and their different patterns of growth and decline. The important psychological distinction in the theory was between process (fluid intelligence) and product (crystallized intelligence) (Cattell, 1963).

The theory of fluid and crystallized ability attracted little attention, possibly because Cattell soon left Harvard for a research professorship at the University of Illinois. There he turned away from the study of human abilities and returned to his earlier research interest of applying the methods of factor analysis to the study of personality. He later wrote, “I had not learned . . . that more original and vital ideas than mine have collected dust on bookshelves for lack of exegesis by their parent or some scholarly leader” (Cattell, 1971, p. x). Twenty years were to elapse before Cattell was to return to the theory of fluid and crystallized abilities with new data. In the 1963 formulation of the theory, Gf was hypothesized to reflect the physiological integrity of the organism useful for adapting to novel situations that, when invested in particular learning experiences, produced Gc. Thus, Gf was now hypothesized to be physiologically determined, whereas Gc was “a product of environmentally varying, experientially determined investments of Gf.” (Cattell, 1963, p. 4)

Although intuitively appealing, the hypothesis that Gf reflects physiological influences and is thus a better measure of the true intelligence of an individual is perhaps the most controversial aspect of the theory. Several prominent theorists accept the fluid–crystallized distinction, and some also subscribe to the investment theory of aptitude. But they do so without assuming that fluid ability represents something more innate than crystallized ability. For example, Cattell’s student and collaborator, Horn (1976), interpreted Gf simply as “facility in reasoning, particu-
larly in figural or non–word symbolic materials” (p. 445). Cronbach (1977) went even further and argued that “fluid ability is itself an achievement” that reflects the “residue of indirect learning from varied experience” (p. 287). More recently, Horn (1985) echoed the same theme: “There are good reasons to believe that Gf is learned as much as Gc, and that Gc is inherited as much as Gf” (p. 289). Gc, said Horn, reflects individual differences in “acculturation learning” whereas Gf reflects individual differences in “casual learning” and “independent thinking” (Horn, 1985, pp. 289–290). Horn and others point out that, if tests of fluid abilities were somehow better estimates of the physiological integrity of the organism and if achievement tests were more a product of experience, then scores on tests of fluid abilities should show relatively higher heritabilities, which they do not (Horn, 1985; Humphreys, 1981; Scarr & Carter-Saltzman, 1982). These theorists also reject using tests of fluid ability as measures of “capacity” or “potential” against which achievement can be gauged (Cronbach, 1977; Humphreys, 1985; R. L. Thorndike, 1963). On the contrary, some argue that fluid abilities are among the most important products of education and experience (Snow & Yalow, 1982).

Recent Changes in Gf-Gc Theory

The most important change in Gf-Gc theory in recent years has been the addition of several other second-order factors to the model. These developments are summarized somewhat differently by Cattell (1971) and by Horn (1985). Horn identified 10 second-order factors: two deep processing factors (Fluid Ability and Crystallized Ability), three perceptual organization factors (Visualization, Clerical Speed, and Auditory Thinking), three associational processing factors (Short-Term Acquisition and Retrieval, Long-Term Storage and Retrieval, and Correct Decision Speed), and two sensory reception factors (Visual Sensory Detection and Auditory Sensory Detection). Figure 1 shows how these factors can be arrayed along a continuum that progresses from surface to deep processing or from infancy to adulthood.

The model is frankly speculative. “I know very little about human abilities,” writes Horn (1985). “All I can do is write articles about them, talk about them, and specify models for them. The more I talk and write and model, the more I realize how little I really know about this complex realm of human functioning” (p. 293). Nevertheless, the model summarizes much of what is known about the organization of human abilities, and it is, in the main, consistent with the abilities Carroll (in press) has thus far identified in his massive review and reanalyses of 60 years of factor-analytic studies of human abilities. Recent research on the four most widely studied broad factors in this model is presented in the next section.

Unpacking Existing Tests and Constructs

Tests of Fluid and Crystallized Abilities

Tests of fluid ability require novel problem solving, much like many of the intelligence tests developed during the first half of the century—particularly the so-called nonverbal or performance tests such as matrices or block design. These tests require subjects to reason with moderately novel figural or symbolic stimuli. For this reason, complex spatial tests often load strongly on the Gf factor (Lohman, 1979). Span tests and other measures of what Jensen (1969) calls Level I ability also often load significantly on the Gf factor (Horn, 1985). Tests of crystallized ability, on the other hand, require the examinee to display an understanding of
concepts and skills taught in some domain, particularly in school. Verbal knowledge and skills are emphasized, although numerical computation and mechanical knowledge tests often load significantly on Gc factors.

Recently the Stanford-Binet was revised along the lines of the theory of fluid and crystallized abilities. The particular version of Gf–Gc theory on which the new Stanford-Binet is based combines the hierarchical model of intelligence of Vernon (Vernon, 1950) and the quasi-hierarchical model of intelligence of Cattell (1963).

**FIGURE 1. A model of ability organization within developmental and information processing hierarchies**

The three-level hierarchy includes a General Reasoning factor, G, at the top. Three broad group factors—Crystallized Abilities, Fluid-Analytic Abilities, and Short-Term Memory—constitute the second level. Three more specific factors make up the third level. G is interpreted "as consisting of the cognitive assembly and control processes that an individual uses to organize adaptive strategies for solving novel problems" (R. L. Thorndike, Hagen, & Sattler, 1986, p. 3). Thus, the authors adopt Snow's (1981) definition of Gf as their definition of G. This is a reasonable equation since the Gf factor is invariably highly (Cattell, 1971; Lohman, 1979), or even perfectly (Gustafsson, 1984), correlated with G. Crystallized abilities are represented by both verbal and quantitative reasoning tasks. These abilities "are greatly influenced by schooling, but they are also developed by more general experiences outside of school" (p. 4). Fluid-analytic abilities are estimated by figural and spatial tasks. Fluid abilities are thought to involve "the flexible reassembly of existing strategies to deal with novel situations." Further, the authors acknowledge that these abilities are also developed, but they are developed from more general experiences than schooling. Finally, the Short-Term Memory factor is represented by tests requiring memory for beads, sentences, digits, or objects. Thus, the new Stanford-Binet attempts to fit old tasks into a more recent theory of intelligence. But do we really understand these tasks well enough to defend the inference that different combinations of them reflect different abilities? What happens when we try to look at the processes subjects use when solving test items or when acquiring the knowledge they sample? In other words, is it possible to develop process theories of abilities?

**Verbal-Crystallized Ability**

Specific verbal processes. Verbal abilities hold a prominent place in all theories of intelligence. It is not surprising, then, that some of the first efforts to understand intelligence in terms of cognitive processes focused on verbal abilities. Hunt and his colleagues have reported several studies of the information-processing characteristics of subjects who differed in verbal-crystallized abilities. Their work is of particular interest because it deals with an important facet of intelligence and because it shows the strengths and weaknesses of both the newer cognitive-experimental approach and the traditional correlational approach to the study of intelligence. The aim of this line of research is aptly summarized in the question, "What does it mean to be high verbal?" which was the title of a report by Hunt, Lunnæberg, and Lewis (1975). The method used in this and several other studies was to select college students with extremely high or low scores on the verbal section of a college entrance examination, to administer to these subjects a battery of presumably well understood experimental tasks, to estimate information-processing scores for each subject on each experimental task, and then to relate these scores to scores on the reference verbal-ability tests using some type of correlational analysis.

For example, in one experimental task, subjects were required to compare pairs of letters of the alphabet, and to respond "yes," if the two letters were physically identical (as in "aa" or "AA"), or "no," if they were different (as in "aA" or "ab"). In a second task, similar pairs of letters were presented, but this time pairs were to be judged according to their names. Thus, in Task 1, the correct answer to the pair "Aa" would be "no," whereas in Task 2, the correct answer would be "yes." An information-processing model for Task 1 (Physical Comparison) would posit proc-
esses for encoding the appearance of the two letters, comparing these representations, and then responding. A model for Task 2 (Name Comparison) would include all of the processes required by Task 1 plus an additional process to retrieve the name codes. Thus, the difference between the time to respond to a given pair of letters in Task 2 and the same pair of letters in Task 1 provides an estimate of the time needed to perform this additional process. The resulting score is called the NIPI (Name Identity minus Physical Identity) difference and has been widely studied as a measure of the speed of accessing overlearned name codes. Correlations between the NIPI score and measures of verbal comprehension are typically about \( r = -0.3 \), suggesting that subjects high in verbal ability access name codes faster than subjects low in verbal ability.

These and other results are consistent with both a hierarchical model of human abilities and with current theories of the way knowledge is represented in memory. In particular, the information-processing tasks used by Hunt et al. (1975) appear to measure specific verbal abilities found in the lower branches of hierarchical models of abilities. Performance on many of these tasks depends on the subject’s ability (a) to produce a rapid, fluent response and/or (b) to remember the order in which information was presented. This latter ability is sometimes represented in models of memory by a special type of memory code called a linear order (Anderson, 1983). Such a code preserves the sequential structure of an event: what came first, then next, then next, and last. Spelling tests require this sort of memory code; one must not only remember the correct letters but also their proper sequence. Similarly, sequencing arbitrary phonemes into words, such as when learning a new language, or sequencing arbitrary words into strings of words, such as when memorizing the names of the letters of the alphabet, days of the week, or lines in a poem, seems to depend in part on the ability to code information in this way.

Research relating scores on experimental tasks to scores on verbal ability tests also has revealed important limitations in efforts to generalize from laboratory tasks to test behavior. First, seemingly simple experimental tasks can measure different abilities in different subjects. For example, Hunt and others (see, e.g., Hunt, Lunneborg, & Lewis, 1975) have used a sentence verification task in which subjects are shown a phrase such as “star above plus” and a picture which either conforms with or contradicts the sentence. Subjects must determine whether the picture and sentence agree. However, minor variations in procedure can substantially alter the way subjects solve this task (Glushko & Cooper, 1978). More importantly, in any given procedure, subjects can differ in the way they solve the task: some create a mental picture from the phrase and compare it with the picture, and some convert the picture to a verbal description and compare that description with the phrase (Macleod, Hunt, & Mathews, 1978).

A second limitation stems from the low correlations between scores representing particular information processes on experimental tasks and scores on reference tests of verbal abilities. Keating and MacLean (1987) argue that the main contribution of the information-processing approach to the analysis of intelligence is that it permits investigators to identify particular mental processes such as rate of rotation or speed of lexical access. The value of the process approach diminishes quickly when these parameters show low correlations with other measures or with similarly labeled parameters derived from other tasks. Keating and MacLean are particularly critical of studies in which Hunt abandoned process parameters and
instead defined latent "process factors" based on correlations among total reaction time (RT) or errors on experimental tasks. Using composite indices in this way, they claim, comes close to "dismissing the logic of the original cognitive correlates approach" (p. 259). Such composite indices cannot be used to "explain" composite indices computed in the same way on ability tests.

Part of the confusion here surely stems from different expectations about what process parameters represent. It is commonly assumed that, by fitting an information-processing model to a task and by decomposing a composite index (total correct or total latency) into component indices, one has also decomposed individual differences on the task into cleaner components. This is not the case. Actually, individual differences in component scores (e.g., rate of rotation) salvage individual differences relegated to the error term when performance for each individual is summarized in a single score such as number of problems solved correctly, or mean response latency. Recapturing variance from the error term might be a profitable activity but only when items on the task show poor internal consistency. Even then, it must be recognized that such scores do not represent a decomposition of the individual differences variance reflected in total or average scores.

Low correlations between scores thought to represent particular verbal processes and reference verbal-ability tests may also mean that much of the knowledge or some of the cognitive processes that account for general crystallized abilities (Gc) as measured by tests are not required by the experimental tasks. Experimental tasks in which subjects are required to infer the meaning of unfamiliar words from context sometimes show much higher correlations than do simple laboratory tasks with both Gc scores and general reasoning scores (Sternberg & Powell, 1983). This suggests that the low correlations obtained by Hunt et al. (1975) may estimate the contribution of specific verbal processes to Gc. Much of the remaining variability in Gc is better attributed to the ability to apply general reasoning skills and prior knowledge to the task of understanding verbal material and learning from it.

Reading comprehension. Nowhere is this interdependence of specific component processes, general reasoning abilities, and prior knowledge better demonstrated than in reading. Reading comprehension is highly correlated with general verbal abilities, particularly in school-age populations. Thus, research on reading comprehension not only illuminates an important aspect of Gc but also shows how diagnostically useful tests can be derived from theory and how studies of individual differences can in turn reveal needed changes in the theory. J. R. Frederiksen's (1982) work is perhaps the best example of this reciprocity. Frederiksen began by developing a general model of reading from his own research and that of many other investigators. He eventually distinguished three types of information-processing skills used in reading: word-analysis processes (e.g., encoding single- and multiletter units, using phonics skills), discourse analysis processes (e.g., retrieving word meanings, resolving problems of reference), and integrative processes (e.g., combining information from pictures and text). Frederiksen then constructed a test battery to measure some of these skills. Measures were validated by using both experimental and correlational techniques. Later, training tasks were devised to assist poor readers in acquiring deficient skills.

Other theories of reading ability have been advanced in recent years. For example, Perfetti (1986) distinguishes three types of component processes in his theory: lexical access, proposition encoding, and text modeling. Lexical access refers to the
process by which word meanings are activated in long-term memory. Individual word meanings are then combined and retained in working memory in predicate-like structures called propositions. These in turn are combined with the reader’s prior schematic knowledge to form a text model. This model, then, represents the reader’s understanding of the text. Kintsch (1986), in another theory of text comprehension, argues that two types of mental models must be coordinated: a text model, which contains the reader’s representation of the propositions embedded in the text, and a situation model, which might be a mental image of the situation described by the text. For example, in following directions to assemble a toy, the text model might represent the ideas implied by the words, “Attach wheel K to spindle Q using two 5/16 washers and a large hex nut.” The situation mental model might be represented by an image of what one is supposed to do. Pictures, illustrations, good description, metaphor, and analogy facilitate the generation of good situation models. A well structured text that follows a familiar schema and uses familiar words facilitates the construction of a coherent text model.

Mental models may be an important link in the individual difference equation as well. A central problem in the definition of verbal abilities has been the overlap between measures of reasoning abilities and measures of verbal comprehension. However, theories of reasoning (Holland, Holyoak, Nisbett, & Thagard, 1987; Johnson-Laird, 1983) also emphasize the construction and the coordination of mental models. Thus, process analyses reveal commonalities between tasks (and the ability constructs they define) not apparent in armchair analyses.

A similar argument may account for the high correlation between reasoning and vocabulary scores. The meaning of an unfamiliar word is usually inferred from the contexts in which the word has been embedded. (Daalen-Kapteijns & Elshout-Mohr, 1981; Marshalek, 1981; Sternberg & Powell, 1983). This process is most successful when the learner generates a good schema (or model or working hypothesis) about the meaning of an unfamiliar word when it is first encountered. This schema can then be confirmed or contradicted by evidence from subsequent contexts. Low-verbal subjects are less likely to use this strategy than are high-verbal subjects. Thus, vocabulary tests that use abstract words (i.e., words whose meanings are difficult to infer from a single context) show higher correlations with reasoning than do vocabulary tests of comparable difficulty composed of infrequent words (Marshalek, 1981).

Spatial-Visualization Ability

Spatial tasks have long been used as psychological tests. Before 1915, Porteus had used such “performance” tasks to estimate the intelligence of linguistically different or disabled examinees. Spearman also originally used such “performance” tests as a measure of g, a tradition he attributes to Itard (1801, cited in Spearman & Wynn Jones, 1950). Spatial tasks also figured prominently in the Army Beta examinations of World War I. However, beginning with Kelley (1928) and then El Koussy (1935), such tasks were studied in their own right, and several specific spatial abilities were identified (Smith, 1964). Nevertheless, spatial or figural reasoning tasks have continued in their role as measures of general abilities, particularly Gf.

As with verbal abilities, cognitive research on spatial abilities may be divided into (a) attempts to develop general theories of spatial thinking that ignore individ-
David F. Lohman

ual differences (e.g., Pinker, 1984; Shepard & Cooper, 1982), and (b) attempts to explain individual differences on existing tests of spatial abilities, either through correlations between scores on spatial tests and performance on laboratory tasks or through the construction of information-processing models for particular spatial tests. In contrast to recent research on verbal abilities, however, only a few studies have examined correlations between scores from laboratory tasks and scores from existing tests. Instead, most effort has been directed toward attempts to build information-processing models that describe how subjects solve particular spatial tests (see, e.g., Pellegrino & Kail, 1982). This is because most spatial tests are process-intensive in the same way that most verbal tests are knowledge-intensive. In other words, although some interesting processing occurs when subjects take a vocabulary test (Sternberg & McNamara, 1985), most of the complex processing occurred at the time the words were learned. Conversely, although spatial knowledge has an important impact on spatial problem solving (Lohman, 1988), whether subjects solve such problems depends heavily on the processes they employ during the test.

Theories of spatial thinking (e.g., Kosslyn, 1980) distinguish two types of spatial knowledge: knowledge best modeled by quasi-pictorial mental representations (e.g., appearance of a particular object) and knowledge best modeled by abstract, proposition-based memory representations (concepts of symmetry, proportionality, closure, etc.). Each type of representation can be transformed by a different class of mental operators or procedural knowledge. Quasi-pictorial representations can be subjected to various analog transformations such as a rotation or synthesis (Shepard & Cooper, 1982). Propositional representations can be subjected to the same general and specific cognitive operators (e.g., means ends analysis) that can be applied to propositional knowledge derived from other sources (e.g., linguistic inputs). Transformations such as rotation, then, are of interest primarily for the constraints they place on the type of mental representation used. Thus, many spatial-ability tests present items which seem to require for their solution analog transformations such as rotation, reflection, transposition, or synthesis.

Research on how subjects solve spatial tests has turned up several surprises. One persistent finding has been that all subjects rarely solve figural tasks in the same way. For example, in a series of experiments on visual comparison processes, Cooper (1982) identified two markedly different strategies. Some subjects appeared to rely on a serial, analytic process to compare forms whereas others relied on a parallel, holistic process. Complex tasks—such as the paper-folding tasks or formboard tasks commonly seen in mental tests—elicit an even wider range of alternative solution methods. Some subjects solve items on such tests by generating mental images that they then transform holistically. These high-spatial subjects excel in generating, retaining, and transforming mental representations that preserve information about the configuration of a figure. They also use their spatial knowledge to decompose unfamiliar visual shapes into simpler, more familiar shapes. Other subjects rely on general reasoning skills or external aids (such as line drawings) to solve problems. Others use still different processes. But most subjects use more than one type of processing, generally shifting from one strategy to another as problems increase in difficulty (Lohman, 1988). Such within-subject variability in solution strategy challenges simple information-processing models of spatial tests. Strategy shifting may partially explain why complex spatial tests are often good
measures of g or Gf. Appropriate flexibility in adapting solution methods to meet personal limitations and changing item demands appears to be a central aspect of any process theory of Gf (Snow & Lohman, 1989).

**Fluid-Reasoning Ability**

There has been considerably more research on reasoning or general fluid ability than on either general crystallized or general visualization abilities. However, attempts to understand how subjects solve Gf tasks such as analogies, classification, and series completion that have ignored differences in processing strategy (by averaging over items) or reduced the need for alternative strategies (by drastically simplifying items) have generally produced experimental tasks that show little relationship with scores on reference Gf tests. Put another way, simple items that are all solved in the same way by all subjects probably require little of what we call intelligence.

The effects of simplifying a complex task so that it could be studied experimentally and ignoring within-person strategy shifts were perhaps most evident in Sternberg's (1977) first investigation of analogical reasoning. Sternberg hypothesized that subjects use several different or “component” processes when solving analogies such as “Up is to down as left is to (a) back (b) right” or A:B::C:D1, D2. According to Sternberg’s theory, subjects (a) first read and understand each term in the analogy (encoding), (b) determine the relationship between the A and B terms (inference), (c) infer the relationship between the A and C terms (mapping), (d) generate an ideal answer by applying the A-B relationship to C (application), and (e) compare their ideal answer with the options provided (comparison). If none of the presented options meet the subjects’ criterion for acceptability, they then recycle through some or all of the preceding steps (justification) and finally choose an option and respond (response). Component processes were assumed to be executed serially. Different models were then formulated by deleting particular processes (e.g., mapping, justification) and by specifying different modes of execution for a given process (e.g., self-terminating or exhaustive). Three important results were obtained. First, models were quite successful in accounting for variabilities in response latencies and, to a lesser extent, in response errors. Second, the data from most subjects were well fitted by a single model, suggesting that most subjects used the same strategy. Third, estimates of speed of executing particular component operations showed small and inconsistent relationships with reference reasoning tests. Unexpectedly, the highest correlations were observed for the preparation-response component. Thus, the componential analysis appeared successful, but those components hypothesized to reflect the essence of reasoning seemed not to measure reasoning at all.

Later studies in which better practiced subjects attempted more complex items did show significant correlations between component scores and scores on reasoning tests (Bethell-Fox, Lohman, & Snow, 1984; Sternberg & Gardner, 1983). It appears that problems must be more than trivially difficult before individual differences in reasoning are observed. Further, items must also vary somewhat in the processing demands they place on examinees. This means that problems must be moderately novel.

Novelty is an ancient theme in the psychology of individual differences. From Stern (1912/1914) to Sternberg (1985), theorists have argued that intelligence is
best displayed when tasks are relatively novel. Cognitive psychologists are only beginning to understand how subjects transfer prior learning to analogous situations (Gick & Holyoak, 1983). The problem, of course, is that what is novel for one person may not be novel for another person or even for the same person at a different time. It appears that inferences about how subjects solve items that require higher level processing must be probabilistic, since the novelty of each item varies for each person.

Snow (1981) has integrated these and other research results in the following hypothesis on the nature of fluid and crystallized abilities.

Gc may represent prior assemblies of performance processes retrieved as a system and applied anew in instructional or other performance situations not unlike those experienced in the past, while Gf may represent new assemblies of performance processes needed for more extreme adaptations to novel situations. The distinction is between long-term assembly for transfer to familiar situation vs. short-term assembly for transfer to unfamiliar situations. Both functions develop through exercise, and perhaps both can be understood as variations on a central production system development. (p. 360)

The point about “exercise” derives from E. L. Thorndike’s (1903) theory of learning whereas the point about “production system” derives from the ACT* model of Anderson (1983), which is discussed later.

Mental Speed

The fourth and last broad factor in Horn’s (1985) model that will be examined here is sometimes called General Speed, sometimes Clerical Speed, or sometimes simply, Mental Speed. There is a new interest in this construct, whatever it is called. However, like most other ability constructs, mental speed has a long history in educational and psychological measurement. E. L. Thorndike, Spearman, and Thurstone all addressed the question of whether mental speed should be distinguished from power (or altitude). For example, although mental speed was one of the four dimensions of his model of intelligence, E. L. Thorndike considered speed less important than altitude (see E. L. Thorndike et al., 1926). On the other hand, Spearman (1927), citing studies which showed high correlations between scores on a time limit test and scores on the same test after an extended period of time, concluded (erroneously) that speed and power (or altitude) were interchangeable. Thurstone (1937) proposed a three-dimensional model that related ability, speed, and motivation. Like E. L. Thorndike, he defined ability in terms of power or altitude in his model (although many of the ability factors he identified in his empirical studies were based on simple, highly speeded tests).

Individual differences in mental speed have been studied in several paradigms, two of which are summarized here. Research in the first paradigm at first sought to estimate the subjects’ “natural” rate of thinking (Hunsicker, 1925). This search led to the identification of several personality factors such as Carefulness, Persistence, and Impulsivity that described subjects’ typical trade-off between speed and accuracy. It also led to the identification of several cognitive speed factors, such as Perceptual Speed, Clerical Speed, and eventually, to claims of a General Speed factor.

Research in the second paradigm, which may be traced back to Galton (1869) has sought to define intelligence as a physiological rather than as a psychological
or sociocultural construct. Thus, the aim is to determine the integrity and efficiency of neurological mechanisms thought to underlie intelligent thought and action. Preferred indicators of intelligence in this paradigm are measures of sensory acuity, speed of detecting a stimulus or discriminating between two stimuli, and, in more recent work, patterns in recordings of electrical activity in the brain. Correlations are then computed between these measures and more global indices of intelligence, such as teacher ratings, course grades, or scores on existing intelligence tests. Work in this paradigm had hardly begun when it was abandoned by most psychologists, partly because of studies like that of Wissler (1901), but perhaps in larger measure because of the success of Binet’s test. Wissler, working under the direction of James McKeen Cattell at Columbia (who had in turn worked with Galton for a short time), found that a measure of RT was uncorrelated with grade point average in a sample of university students. The RT paradigm has recently been revived by Jensen, Eysenck, and others.

Speed factors. Variation in the relative emphasis tests placed on speed or level of performance is an important confound in much of the literature on human abilities. The primary factors identified by Thurstone and his followers, particularly Guilford, were often defined by tests that contained simple, similar, highly speeded items. Complex versions of the same tests administered under conditions which emphasize level or altitude invariably show stronger loadings on the general factor and little evidence of the fractionalization of ability that occurs when simple, speeded tests are administered (Lohman, 1979). This is because individual differences in the speed with which subjects can solve relatively simple problems in a domain show only weak correlation with the complexity of a problem of the same type which subjects can solve when time is not a factor (Horn, 1985; Kyllonen, 1985).

The question remains, though, whether some or all of these various speed primaries may define a higher order or General Speed factor. Although several investigators have claimed to have identified a General Speed factor, closer examination shows that such factors are often little more than overblown Clerical Speed or Perceptual Speed factors. General differences in speed of processing may well exist, but they are difficult to identify by factor analyzing speed scores from a battery of tests. The major reasons are that one cannot make unambiguous comparisons of response latencies across individuals unless (a) all subjects correctly solve all items, (b) all subjects adopt the same trade-off between speed and accuracy, and (c) neither of these factors vary systematically across tasks. One way to avoid these problems would be to use a single task that is so simple that everyone can solve it and that is not much influenced by the individual’s decision to emphasize speed or accuracy. Recent studies of reaction time aim to fit both of these criteria.

Recent research on reaction time. The primary dependent measure in much cognitive research is response latency, usually on simple tasks. Those who study individual differences raised the question of whether individual differences in latencies on these laboratory tasks would show any relationship with individual differences on other tasks that presumably required the same processes (Underwood, 1975) or with ability variables commonly assessed by mental tests (Hunt et al., 1973; Snow et al., 1976). But the main goal of researchers like Hunt, Snow, and Sternberg was to develop and test information-processing models of theoretically interesting cognitive tasks or of tests commonly used to estimate important ability constructs, not to propose new measures of mental speed. However, this
was precisely the goal of another group of researchers. Led by Jensen in the United States and Eysenck in the United Kingdom, these researchers saw possibilities for new measures of intelligence in response latencies on simple tasks and other indices of cognitive efficiency presumably unaffected by intention or experience.

Jensen's work. Jensen sparked new interest in the relationship between RT and G (intelligence) by showing significant correlations between choice (or discrimination) RT and measures of G. Jensen's work has generated much discussion. In part this is because his goal seems to be to isolate a culture-free measure of intelligence. Individual and group differences on such a measure could then not be interpreted "as reflecting only differences in cognitive contents and skills that persons have chanced to learn in school or acquire in a cultured home" (Jensen, 1980, p. 704).

The apparatus Jensen has used in his studies contains a center "home button" surrounded by 8 light/button pairs. Different light/button pairs can be covered to manipulate the number of stimulus–response pairs between 1 and 8. The task is to hold a finger on the home button until one of the exposed lights is activated and then turn it off as quickly as possible by moving the finger from the home button to the button directly below the activated light. Two time intervals are recorded: (a) the time between the onset of the stimulus light and the release of the home button (called RT), and (b) the additional time required to move the finger to the button below the activated light and press it (called movement time). In a typical experiment, subjects receive a few practice trials, followed by 15 trials at each of four levels of task complexity: 1, 2, 4, or 8 light/button pairs exposed. Typically, RT increases linearly with the log of the number of buttons exposed. Jensen found that the slope of this function, which is taken as an estimate of the rate at which a person processes a single unit of information, and G correlated negatively, with \( r = -\cdot41 \) being the most often cited correlation. In addition, the correlation between RT and G increases as task complexity is increased from 1 to 8 light/button pairs, suggesting that the greater the information-processing burden, the greater the demand on G.

Jensen's work has been praised by some (e.g., Eysenck, 1982) and criticized by others (e.g., Longstreth, 1984; Carroll, 1987). In particular, Jensen's claim that performance on the choice RT task is not influenced by practice, motivation, or instructions to alter speed–accuracy trade-off has been questioned (Carroll, 1987; Longstreth, 1984). Longstreth also raises a number of fundamental questions about Jensen's procedure, such as the routine confounding of practice with task complexity. Carroll questions the replicability and interpretation of Jensen's results. He suggests that differences between individuals in average RT may better be described as differences in the variability in RT for a given person over trials. This is because RTs have a lower limit, and thus individuals with more variable RTs would tend to have higher mean RTs because they are more likely to deviate upward from the lower limit. This suggests that the observed correlation between RT and G may in part reflect differences in attentional control and not simply differences in the speed of neural conduction or the rate of neural oscillation, as Jensen hypothesizes.

Attempts to replicate Jensen's findings usually find some relationship between RT and G (most often between the variability of RTs for individual subjects and G, with lower G subjects having more variable RTs). But replications consistently fail to find that low G subjects show greater increases in RT with increases in the
number of exposed light/button pairs than do high G subjects (Barrett, Eysenck, & Lucking, 1986; Carlson, C. M. Jensen, & Widaman, 1983; Jensen, 1987).

Although controversy about Jensen’s work continues, there is some consensus on the main findings. First, correlation between G and RT is generally somewhat lower for the simple RT condition (one light/button pair exposed) than for the discriminative RT conditions (two or more light/button pairs exposed). Second, correlations between discrimination RT and G vary widely. However, replicable correlations are generally in the −.2 to −.4 range. Conditions with more light/button pairs (e.g., 8) do not yield dependably higher correlations with G than conditions with fewer light/button pairs (e.g. 2). Indeed, it is a common finding that correlations between RT and G decline as more and more complex information processing is required. More complex tasks allow multiple strategies and are prone to differences in the speed–accuracy trade-off subjects adopt. Third, the variability in RT over trials often correlates as highly with G as does mean or median RT. Thus, attention control (or, conversely, distractibility) may be as important as speed of processing in this task. Fourth, Jensen’s claim that RT increases linearly with the log of the number of exposed light/button pairs has been repeatedly confirmed. However, other investigators have not been able to confirm his claim that individual differences in the slope of this line correlate with G. It is unclear whether this is due to persistent methodological inadequacies in these studies (which usually follow Jensen’s procedures), as Longstreth (1984) notes, or whether this reflects a more fundamental error in Jensen’s theory, as Eysenck (1987b) now claims.

Eysenck’s work. Eysenck (1982; 1988) has proposed a theory of intelligence with an even stronger physiological flavor. Following Hebb (1949), Eysenck (1988) distinguished among biological intelligence, psychometric intelligence, and social intelligence. Biological intelligence “refers to the structure of the human brain, its physiology, biochemistry, and genetics which are responsible for the possibility of intelligent action” (p. 3). Eysenck considers biological intelligence to be the purest, most fundamental intelligence because it is “least adulterated by social factors.” He claims it can be measured by the electroencephalogram (EEG), evoked potentials, galvanic skin responses, and perhaps reaction times.

Psychometric intelligence is defined as that intelligence which is measured by psychometric tests. In addition to the core of biological intelligence, is determined by cultural factors, education, family upbringing, and socioeconomic status. However, since only a fraction of the variance in psychometric intelligence (i.e., IQ) can be attributed to genetic factors (Eysenck estimates 70%), IQ should not be confused with biological intelligence.

Social intelligence reflects the ability to solve problems an individual encounters in life. But since so many noncognitive factors are reflected in such performances, Eysenck (1988) argues that “social intelligence is far too inclusive a concept to have any kind of scientific meaning” (p. 45). Thus, for Eysenck, intelligence is a concept that is best studied at the physiological (or even neurological) level, only indirectly represented in intelligence tests, and obscured almost entirely in performances in the real world. This is an extreme view and is not widely shared, at least not by American academics.

As with Jensen’s work, much of the controversy surrounding Eysenck’s work has
centered not so much on the finding of significant correlations between G and EEGs, cortical evoked potentials, and other physiological indices but on the reported magnitude of the correlations. For example, Eysenck's colleague, Hendrickson (1982), reported a correlation of \( r = .83 \) between a measure of evoked potentials and Wechsler IQ for a sample of 219 15-year old children. In 1984, Eysenck claimed that "several replications...have shown the results are essentially reproducible" and that these results were "a most important validation of Galton's concept" of intelligence (published in Eysenck, 1987a, p. 359). However, by 1988, presumably on the basis of new evidence, Eysenck had changed his mind. "It seems unlikely that the correlation between IQ and a physiological measurement of biological intelligence...can exceed the square root of the heritability of IQ," and thus correlations such as those obtained by Hendrickson (1982) are "inherently improbable and unlikely to be replicated" (Eysenck, 1988, p. 12).

**Inspection time.** A similar history attends the reports on correlations between inspection time and IQ. Inspection time is the minimum duration for which two different stimuli must be presented if they are to be perceived as different. Nettlebeck and Lally (1976) reported a correlation of \( r = -.92 \) between the Wechsler Adult Intelligence Scale performance scale and inspection time, but, for a sample of only 10 subjects, 2 of which were retarded. The magnitude of the reported correlations gradually declined as larger and less wide-ranging samples were tested. By 1984, Irwin reported correlations of \( r = -.32 \) and \( r = -.09 \) for auditory and visual inspection times with a verbal intelligence test and correlations of \( r = -.23 \) and \( r = -.27 \) for those same inspection times with a nonverbal intelligence test for a sample of 50 12-year-old children.

In the meantime, Nettlebeck and Kirby (1983) had gathered new data on a large sample of adults and had reanalyzed data from one of their earlier studies. This time they found no correlation between G and slope in the Jensen task and a weak correlation between inspection time and G \( (r = -.3) \) when retarded subjects were excluded. They therefore concluded that their earlier correlations had been inflated by the inclusion of retarded subjects, who were "markedly less efficient" (p. 39) on these tasks. Their conclusions run completely counter to earlier claims:

This outcome raised doubt about the validity of combining data from retarded and nonretarded subjects. Our results run counter to claims that tasks of the kind used [in this study] are largely uninfluenced by cognitive variables [such as strategy], so that findings are not necessarily explained satisfactorily in terms of a mental speed factor. These measures of timed performance do not, at this time, provide a basis from which a reliable, culture-fair measure of intelligence might be devised. (p. 39)

**Summary.** Critics of studies that report correlations between measures such as RT, inspection time, evoked potentials, and G cynically argue that the best predictor of the correlation obtained is the date of the study. The first correlation reported is usually strikingly high, but then the magnitude of the reported correlation declines almost linearly with year of publication, eventually stabilizing on a value in the \(-.1 \) to \(-.4 \) range. Such correlations are theoretically interesting, but they do not justify attempts to replace existing intelligence tests with RT measures, or interpretations of G as a purely physiological phenomena.

One need not descend to the level of neurons to find a plausible account of the role of mental speed in models of intelligence. For example, the rate at which
activation spreads through regions of memory, the rate at which an activated memory loses its activation, and the level of activation needed to allow further processing are all important constructs in modern theories of memory (Anderson, 1983). Direct study of these variables would seem more useful than the study of isolated tasks that have not been designed to estimate specific cognitive processes. Even then, variables thought to reflect the physiological action of the cortex are useful only to the extent that they predict individual differences in behavior labeled "intelligent" in the culture. E. L. Thorndike saw this clearly:

Psychologists would of course assume that differences in intelligence are due to differences histological or physiological, or both, and would expect these physical bases of intelligence to be measurable .... [However], even if one aimed at discovering the physiological basis of intellect and measuring it in physiological units, one would have to begin by measuring the intellectual products produced by it. For our only means of discovering physiological bases is search for the physiological factors which correspond to intellectual production. (E. L. Thorndike et al., 1926, p. 12)

Individual differences in mental speed have an important impact on all of cognition. But neither theory nor empirical evidence justifies attempts to define G in terms of speed, while ignoring the larger contributions of level or altitude in both process and knowledge to this construct we call intelligence.

Attempts to Move Beyond Existing Tests

It has long been recognized that theories of human intelligence have been limited by the selection of tasks included in particular intelligence tests or in factor-analytic studies of abilities. Several theorists (e.g., Cattell, 1971; Guilford, 1959) have proposed schemes for defining the universe of intelligent behaviors, cognitive functions or tasks. The framework can then be used to select or construct tests of different facets of intelligence. In this section, I briefly survey two rational models of this sort: Guilford's (1959, 1985) structure of the intellect (SOI) model and Sternberg's (1985) triarchic theory of intelligence.

Guilford's SOI Model

As director of the Aviation Psychology Research Unit during World War II, Guilford saw the number of factorially defined abilities grow as tests were developed to measure abilities hypothesized to be important in the training and performance of air crews. After the war, he continued to investigate new abilities in his APTitudes Research Project at the University of Southern California. By the mid-1950s, approximately 40 ability factors had been identified in one or both of these efforts (Guilford, 1985). In searching for a way to organize these factors and guide the search for new abilities, Guilford hit upon the idea of grouping abilities by a three-way classification: by the kind of mental process required, by the kind of information processed, and by the mental products generated. The combination of five types of mental processes, four types of content, and six types of product defined the 120 abilities in the structure of the intellect model.10

Although the model has generated considerable research, it has declined in influence in recent years. Questions have been raised about the factor-analytic methods used to identify factors (Horn & Knapp, 1973), about the seeming
fractionation of ability (McNemar, 1964, called the scheme “scatter-brained”), and about the adequacy of the SOI model itself. Some of these challenges have been countered. Elshout, van Hemert, and van Hemert (1975) showed that Guilford's procrastination factor-analytic methods were not as bad as Horn and Knapp (1973) had claimed. Following Humphreys' (1962) suggestion, Guilford (1985) countered criticisms of fractionation by agreeing that higher order abilities may be defined by averaging over cells within the SOI model. In addition, he countered objections that the model did not include auditory abilities by adding another level to the content facet for auditory abilities—raising the total number of cells in the model from 120 to 150. Nevertheless, levels of facets have no convincing foundation other than rational appeal, and the entire product dimension remains poorly validated (Cronbach & Snow, 1977). Excepting the addition of 30 new auditory abilities, over 20 years of research has produced no substantive changes in the model. Perhaps this is because research sought to demonstrate the validity of the model rather than to identify and correct its weaknesses.

Triarchic Theory

Overview of the theory. Sternberg's (1985) theory of intelligence contains three subtheories: a contextual subtheory, an experiential subtheory, and a componential subtheory. The contextual subtheory attempts to specify those behaviors that would be considered intelligent in a particular culture. Sternberg argues that, in any culture, contextually intelligent behavior involves purposeful adaptation to the present environment, selection of an optimal environment, or shaping of the present environment to fit better one's skills, interests, and values. The nature of the adaptation, selection, or shaping can vary importantly across cultures. For example, navigational skills, hunting skills, and academic skills are highly valued as markers of intelligence in different cultures.

However, even if a particular task is thought to require intelligence, contextually appropriate behavior is not equally “intelligent” at all points along the continuum of experience with that class of tasks. According to the experiential subtheory, intelligence is best demonstrated when the task or situation is relatively novel or when learners are practicing their responses to the task so that they can respond automatically and effortlessly. Although many have suggested that tasks must be moderately novel to measure intelligence, Sternberg's theory is unique in its claim that the ability to automatize processing is also a good indicator of intelligence. To date, no convincing evidence has been advanced to support this hypothesis.

In the componential subtheory, Sternberg attempts to specify the cognitive structures and processes that underlie all intelligent behavior. Contextually appropriate behavior at relevant points in the experiential continuum is said to be intelligent to the extent to which it involves certain types of processes. Three types of processes are hypothesized: metacomponents, which control processing and enable one to monitor and evaluate it; performance components, which execute plans assembled by the metacomponents; and knowledge acquisition components, which selectively encode and combine new information and selectively compare new information to old information.

Thus, Sternberg's contextual subtheory describes what types of tasks, situations, and behaviors might be considered intelligent. It is relativistic with respect to individuals and to the sociocultural settings in which they live. In the United States,
the prevailing contextual theory of intelligence involves problem-solving, or fluid abilities; knowledge-based, or crystallized abilities; and social and practical abilities. The experiential subtheory claims that intelligence is relative to each individual’s experience with the task or situation. Only the componential subtheory claims to describe the mechanisms of thought that would be used in any intelligent act.

Evaluation of the Triarchic Theory. Some argue that intelligence as measured in the tradition of Binet and Wechsler is best construed as scholastic aptitude. This tendency to narrow the scope of intelligence tests has been countered repeatedly by those who would extend measurement to domains such as social intelligence (E. L. Thorndike, 1920), creativity (Guilford, 1959), or musical ability (Gardner, 1983) that are sampled inadequately or not at all by existing tests. Those who would extend the purview of existing tests tend to view intelligence as an adjective rather than a noun and argue that tests of intelligence should sample all domains of activity that are valued as intelligent in the culture. Sometimes these unmeasured abilities are essential features of the theorist’s implicit theory of intelligence or that of a larger social group.

Those who view intelligence as a noun usually equate intelligence with individual differences in a particular type of cognition, such as “eduction of relations and correlates” (Spearman, 1927) or “judgment” (Binet & Simon, 1905). However, others view the noun as a shorthand expression for all individual differences in cognition and argue that a good test of intelligence presupposes a good theory of cognition (Hunt, 1986) or at least a good sample of “the repertoire of intellectual skills and knowledge available to the person at a particular point in time” (Humphreys, 1986, p. 98). Sternberg’s triarchic theory attempts to satisfy both of these demands. His contextual theory recognizes the cultural relativity implied when intelligence is treated as an adjective, and his componential theory “[covers] most if not all of the territory of cognitive psychology” (Carroll, 1986, p. 325).

Reactions to Sternberg’s theory have been mixed. Some argue that his triarchic theory is not a theory at all but a “conceptualization” of intelligence (Humphreys, 1984). Sternberg’s theory for testing implies that one should model individual performance on cognitive tasks that represent fluid and crystallized abilities, so that component scores and solution strategy may be estimated for the individual; recognize that comparisons of individuals and especially of groups may be misleading, because tasks are differentially novel or practiced for different individuals and groups; and broaden the sample of tasks included on intelligence tests to better represent skills in adapting to the environment, shaping the environment, or selecting new environments. Here, Sternberg (1985) sees a special need for tests that measure “real-world” or practical intelligence. In several studies, questionnaires designed to assess respondents’ tacit knowledge about managing self, others, and career have shown moderate correlations with various objective criteria of success in the domain (Wagner & Sternberg, 1986). Cronbach (1986) agrees that this is a worthwhile goal for measurement, but he is unimpressed with the verbal tests of practical intelligence Sternberg has thus far developed. He claims that Sternberg’s tests are “quizzes on gamesmanship” (p. 24). Sternberg counters that scores on his questionnaires are generally uncorrelated with measures of verbal intelligence.

Perhaps Ford’s (1986) research on the measurement of social intelligence can provide some useful cues for the measurement of practical intelligence. He argues that better measures can be obtained when social intelligence is defined in terms of
outcomes (i.e., social competencies) rather than in terms of social cognition (e.g., understanding verbal or pictorial displays of social events). However, practical and social intelligence differ in several respects, and each has its roots in a different tradition. Whereas research on social intelligence stemmed from the observation that academic intelligence was no guarantee of social competence, research on practical intelligence began with the observation that academic intelligence was also no guarantee of "common sense." Thus, studies of social intelligence are rooted in the research on social judgments, whereas studies of practical intelligence developed from research on "tacit" knowledge—that is, knowledge that is not explicitly taught or discussed but that may facilitate performance or even be necessary for success in some domain.

Whether or not Sternberg succeeds in his efforts to develop new measures of practical intelligence or better measures of other aspects of intelligence, he has clearly succeeded in unifying diverse—even antagonistic—traditions in research on intelligence.

With his prolific research, writing, and editing activities, Robert Sternberg has probably done more than any other contemporary psychologist to bring back into attention fundamental questions about intelligence—what it is, how it can best be observed and measured, and how it relates to other domains of behavior. (Carroll, 1986, p. 325)

**Integrative Theories in Cognitive Science**

All of the research efforts described to this point have involved the study of individual differences, either in existing tests of intelligence or achievement, or in tasks taken from the laboratories of experimental cognitive psychologists. However, there is an obvious circularity in attempts to understand the nature of intelligence by studying existing tests of intelligence or by identifying the information-processing characteristics of people who have been labeled high or low ability because of their scores on existing tests. Attempts to specify the cognitive character of the target behaviors or achievements such tests aim to predict expand the circle significantly but do not remove the circularity. What is needed is a general theory of human cognition. Measurements of individual differences could then be derived from this theory rather than in a theoretical vacuum. There have been several attempts to put theory before assessment, particularly in the measurement of reading disabilities (Frederiksen, 1982) and (less successfully) in the measurement of spatial abilities (Poltrick & Brown, 1984). But the term intelligence connotes a much broader effort.

A central question in cognitive science is whether human cognition is best modeled as a unitary system or as a collection of independent systems or modules. This debate parallels the Spearman-Thorndike/Thurstone controversy over g versus multiple factors in differential psychology (see R. M. Thorndike & Lohman, 1989). Much early theorizing presumed a unitary system, as Newell and Simon (1972) advocated in their General Problem Solver. This program aimed to solve a broad array of reasoning problems using general heuristics. By the late 1970s, however, the pendulum was beginning to swing the other way. Led by Chomsky (1980) and Fodor (1981), a modular view of cognition gained popularity. Modularists argue that the mind is best construed as a collection of independent information-
processing systems, including systems for language, visual processing, music, and other specialized mental contents. Chomsky even describes such faculties as "mental organs," analogous to physical organs such as the heart. Modularists point to findings from neuropsychology on apparent localization of different mental functions in different regions of the brain and to factor-analytic and other studies of individual differences which show that musical, spatial, numerical, and other abilities can be distinguished (Gardner, 1983). Most modularists deny the need for a central or executive processor. Some recognize these higher thought processes but argue that cognitive science cannot explain them (Fodor, 1981). [Modularists recognize higher thought processes, but they deny the need for a central or executive processor and argue that cognitive science also cannot explain them (Fodor, 1981).]

Anderson's ACT* Theory

Several research efforts, most notably that of Anderson and his colleagues, have opposed this side of modularity. In a series of monographs (Anderson & Bower, 1973; Anderson, 1976, 1983), Anderson has developed and refined his Adaptive Control of Thought (ACT) system, culminating in the latest version, ACT*. The system is too complex to describe more than its general features here. (The reader is referred to Chap. 1 of Anderson, 1983.)

First, Anderson (1983) claims that all "higher cognitive processes, such as memory, language, problem solving, imagery, deduction, and induction, are different manifestations of the same underlying system" (p. 1). Nevertheless, ACT* posits special-purpose "peripheral systems" that convert information presented to the senses into distinctive perception-based memory representations or codes, such as images (that preserve information about configuration) and temporal strings (that preserve information about temporal order). Other perception-based memory codes (e.g., olfactory, kinesthetic) seem likely, but they have not been much studied. The peripheral systems that create and process these perception-based codes function like the modules Fodor posits.

Higher cognitive processes, however, are thought to depend more heavily on a different type of memory representation that preserves the meaning of an event. Indeed, Anderson (1983, 1985) argues that this type of abstract code dominates long-term memory, even for memories that might appear to be more perception based. For example, much of what we remember about a visual scene depends on our interpretation and understanding of the visual display. On this view, meaning-based representations (such as the idea of roundness) are derived from particular perception-based memories (such as memories for many particular round objects).

This multicode theory of memory has several interesting analogs in research on individual differences. For example, specific learning disabilities may be caused by a dysfunction in one or more peripheral systems that encode information from the environment into memory or decode the products of thinking into particular responses. Conversely, the dominance of the meaning-based code in human cognition corresponds to the dominance of the general factor in individual differences on complex tasks that seemingly emphasize different mental contents or processes. Indeed, general ability—as typically estimated—may reflect the ability to create, transform, and retain meaning-based mental representations (Snow & Lohman, 1989).
A second feature of Anderson’s ACT* theory that can inform theorizing about intelligence is the distinction between declarative and procedural knowledge. These two types of knowledge are posited in many, although certainly not all, AI theories. Declarative knowledge is knowing that something is the case. Procedural knowledge is knowing how to do something.\textsuperscript{11} Declarative knowledge is represented by a network in which nodes are like idea units, and procedural knowledge is represented by conditional imperatives of the form, “If a certain condition holds, then perform a certain action.” Thus, procedural knowledge is dynamic; declarative is static. Procedural knowledge can be executed automatically, even unconsciously; declarative knowledge is often accessed slowly and consciously. Each is also acquired with different proficiency and by different methods. On one hand, new declarative knowledge can be acquired relatively quickly (often in a single trial), often by elaborating relationships with previously acquired knowledge. On the other hand, proceduralization generally requires more extensive practice.

The declarative-procedural distinction has several implications for a theory of intelligence. First, cognitive skills are modeled as forms of procedural knowledge in ACT*. Therefore, those parts of the theory which describe how declarative knowledge is converted to procedural knowledge also describe an important aspect of ability development. Second, the theory predicts the gradual differentiation of abilities some have hypothesized (Garrett, 1946; Anastasi, 1970), and it can explain how the same task (e.g., division) can require general problem-solving skills for the inexperienced examinee and specific problem-solving skills for the more experienced examinee. Third, attempts to measure declarative and procedural knowledge suggest new ways to separate students’ factual knowledge in a domain from their ability to solve unfamiliar problems in the domain. This is an old (Lindquist, 1948) but seldom attained goal in educational measurement. Attempts to assess declarative knowledge usually involve the construction of a map of the examinee’s factual knowledge base. Attempts to assess procedural knowledge emphasize speed of solving problems, methods of classifying them, or errors made in such processing.

Kyllonen and Christal (1989a) have shown that Anderson’s theory can be used as a general framework for the assessment of individual differences. They argue that individual differences on a wide variety of cognitive tasks arise from differences in four primary sources: cognitive processing speed, working memory capacity, breadth and pattern of declarative knowledge, and breadth and pattern of procedural knowledge. Working memory occupies a central position in ACT* and in applications of the four-sources framework to problems of skill acquisition (Woltz, 1988) and reasoning abilities (Kyllonen & Christal, 1989b). For example, in a series of studies, Kyllonen and Christal (1989b) found strikingly high correlations between theory-based measures of working memory capacity and traditional measures of reasoning ability (or Gf). While acknowledging that such correlations are open to multiple interpretations, they argue that individual differences in working memory capacity cause individual differences in reasoning. One interesting implication of this view is that attempts to localize reasoning ability in a particular component process (e.g., inference) are bound to fail since working memory capacity affects success across all component stages of reasoning tasks.

These are but a few of the sorts of connections that can be made between a general theory of cognition and concepts familiar in measurement, particularly educational measurement. Some of these hypotheses may prove useful; others will
surely be discarded. Nevertheless, it would appear that any good theory of intelligence must distinguish between higher level cognitive representations and the processes that operate on them and lower level representations and the processes that mediate between the world and the individual. Such a differentiation may take the form of a hierarchical system: a base of built-in, primitive mechanisms that operate in parallel with processes not accessible to introspection and a second level of processing that is serial, often open to introspection, and can be modified with some flexibility (Gardner, 1985). A good theory of intelligence must also acknowledge the crucial role of knowledge in all of cognition (Glaser, 1984). A major implication of Anderson's theory of research on skill acquisition and of research on expertise is that aspects of thinking that were once considered elementary, wired-in processes are now understood to be knowledge that has been automatized ("compiled" or "proceduralized") through practice. Thus, understanding abilities means understanding individual differences in learning and development.

**The "New Connectionism"**

Critics of the computer metaphor for human thought have long pointed to the discrepancy between the serial, digital "Von Neumann" computer and the parallel, analog nature of much human thought. Cognitive psychologists countered that it was often impossible to distinguish between a serial model, in which one stage of processing follows the heels of another, and a parallel model for the same task, in which all processes start at the same time, run in parallel, but finish at different times. Further, it was argued that parallel processing could be simulated—albeit clumsily—on a serial computer.

These arguments began to lose their appeal as parallel processing computers were constructed and as deliberate efforts were made to make computational models of thought conform better to biological theories of brain function. This new breed of neurally inspired models of cognition is best exemplified in the work of Rumelhart, McClelland, and the PDP Research Group (1986) and their Parallel Distributed Processing (PDP) approach. Instead of a series of operations on symbols, a PDP model contains thousands of connections among hundreds of cognitive units. Excitations or inhibitions are signaled from one unit to another until the network momentarily achieves a stable state. "Thinking" or "action" occurs as strengths of the connections among units are momentarily altered. Memory is thus modeled as the set of relationships among aspects of events encoded in groups. The pattern of connections and their strengths allow particular "memories" to be recreated when the network is activated.

The PDP approach signals a significant shift from purely serial models of thinking to parallel models. Some have already suggested that a comprehensive account of thinking will require both types of processing—for example, a richly interconnected hierarchy with parallel-processing modules at the base that are dedicated to particular sensory inputs or response systems and a serial, limited capacity system at the apex to model higher order thinking (Gardner, 1985). Such a system mirrors the sort of hierarchical model of human intelligence advocated in various guises by Spearman, Burt, Vernon, and Cattell.

The PDP approach also reflects a shift from theories rich in process but short on knowledge to theories that are rich in knowledge but short on process. There has been a gradual realization in all of cognitive science of the importance of an
extensive, accessible, and well-organized knowledge base for intelligent performance. In AI, early efforts to avoid knowledge in the interest of simplification only served to make the task of modeling human reasoning "harder than it needed to be" (Dehn & Schank, 1982, p. 373). Similarly, there has been a gradual shift in cognitive psychology from the sort of knowledge-free information-processing models that can be neatly summarized in a flow chart to the study of the role of prior knowledge represented as scripts (Schank & Abelson, 1977), schema (Rumelhart & Ortony, 1977), mental models (Johnson-Laird, 1983), and belief systems (Carey, 1986). The importance of knowledge has even emerged in process-intensive tasks, such as those used to estimate spatial abilities. Further, many functions formerly represented as wired-in processes in information-processing models are now seen as acquired proficiencies (i.e., procedural knowledge). Indeed, the goal of measuring knowledge- or experience-free cognitive processes may be a measurement pipe dream, as E. L. Thorndike et al. (1926) suggested. In a way, this newfound role of knowledge in cognitive science parallels the gradual realization by differential psychologists that intelligence is not an innate characteristic of the person but an acquired set of competencies (Anastasi, 1986; Cronbach, 1984).

Future Directions

Including Affect

Kant popularized Aristotle's threefold categorization of mental faculties: cognitive, affective, and conative (knowing, feeling, and willing). By this account, a complete theory of mind must explain not only the cognitive dimension but also the emotional and intentional dimensions as well. Attempts to simply the task of understanding intelligence by ignoring emotion and intention may prove as ineffective as early attempts to ignore knowledge in AI. Indeed, theorists are once again beginning to argue that affect must be included in accounts of learning and cognition (Snow & Farr, 1987). Thus, one direction research on intelligence seems to be taking is to expand its horizons to include affective dimensions long recognized as central to intelligence (e.g. Wechsler, 1939) but rarely combined with the systematic study of the cognitive dimensions (see Royce, 1979, however, for one effort). A theory of intelligence thereby becomes more than an account of human cognition. It becomes an account of affect and perhaps even volition as well. Even when intelligence is treated as a noun, its purview knows no bounds.

From Crystallized to Fluid

A second trend in research on intelligence is moving in the opposite direction. Binet's test was originally designed to predict performance in school. Whatever larger purposes he might have hoped the test might serve, or that others have actually used tests for, it is clear that intelligence tests have always been most heavily used as measures of scholastic aptitude. Researchers have begun to uncover the reasons why such tests predict success in conventional forms of schooling as they have begun to understand the nature of the knowledge and thinking skills that are required by school-learning tasks that are also estimated by intelligence tests. Items on intelligence tests often appear to differ markedly from the sort of school-learning tasks they predict. For example, matrix completion problems and/or paper folding problems do not appear to have much in common with understanding a
story or solving an algebra word problem. Yet intensive analyses reveal a commonality in the processes students use to solve both test problems and school-learning tasks (Snow & Lohman, 1984).

Analyses of existing intelligence tests and of the school-learning tasks such tests were originally designed to predict will continue to be important activities in measurement and instructional psychology. However, the study of school-learning tasks is now viewed by some as the research activity most likely to produce useful results (Cronbach, 1984, p. 300). In fact, there has been a subtle shift in recent years from the study of intelligence to the study of achievement, particularly the acquisition, organization, and use of knowledge in particular domains such as science, mathematics, and literature (Glaser, Lesgold, & Lajoie, 1987). Thus, somewhat paradoxically, new developments in the measurement of intelligence—particularly the sort of intelligence required by and developed through formal schooling—may well come about more through the careful study of achievement than through continued scrutiny of tasks modeled after existing intelligence tests. And there are reasons to be optimistic that such research may produce intelligence tests that are useful for instruction in more ways than are existing tests.

This possibility can be better understood if intelligence and achievement are viewed as points on a continuum of transfer or novelty rather than as qualitatively distinct constructs. Figure 2 shows one such continuum. The horizontal line symbolizes the amount of transfer required by the test or the average novelty of the problems for the typical examinee. At the far left, problems on the test duplicate those taught. As one moves to the right, problems become increasingly novel and require increasing transfer. For example, if students have learned to add numbers in columns, then one could present these same addition facts in column format to require minimum transfer. Presenting the same facts horizontally would require a bit of transfer; embedding the problems in a sentence would require more transfer; and embedding them in a matrix problem in which the rule is “add row 1 to row 2” requires even more transfer. Perhaps creating the matrix items in the first place requires the most transfer. As this example demonstrates, the continuum of novelty in Figure 2 is not limited to general ability but can apply to narrower ability constructs as well. It also illustrates the principle that the same task can elicit different processes from different people, depending on their prior experience.

Important educational objectives may be identified all along this line (Elshout, 1987). Students must learn specific skills, but they must also learn to transfer their learnings to unfamiliar situations and to be creative. Unfortunately, measurement problems increase as one moves from left to right on this scale. Tests that sample no more than those facts and skills explicitly taught are relatively easy to defend.
especially when only limited inferences are made from test scores. Tests that require transfer are more difficult to defend because problem novelty varies from individual to individual and because such tests are usually constructed in ways that encourage grander inferences. Some argue that defensible tests of insight (on the far right) are nonexistent.

Much of the research on intelligence and intelligence tests conducted by Sternberg, Snow, Hunt, Pellegrino and others during the 1970s could be seen as an effort to start in the middle of Figure 2 and move to the left. Both Snow (1978) and Glaser (1976) argued that the ultimate goal of their research on intelligence was to discover how the thinking skills required by such tests are also required for learning in schools. Although much has been learned from these efforts, dependable methods for encouraging the development of fluid abilities have not been discovered, even though many recommendations have been made (e.g., Wagner & Sternberg, 1984). In part, this may be an inevitable consequence of studying tests that were designed to work rather than to reflect a particular theory of cognition. A more fruitful avenue, for education at least, might be to begin somewhere near the left of Figure 2 and work toward the right. Perhaps then educators might finally learn what to teach the so-called “overachiever,” who scores higher on tests of crystallized abilities than on tests of fluid abilities. The recent work of Brown and Ferrera (1985) in estimating a student’s “zone of proximal development” exemplifies one effort toward this goal.

Process Sensitive Tasks

A third trend in research on intelligence is a renewed emphasis on the contextual foundation of the concept “intelligence” in the culture and life-history of the individual. In part, this represents a rediscovery of the fact that, as E. L. Thorndike et al. (1926) put it, “measurements of intelligence rest on judgements of value” (p. 12). But it also represents a breaking down of artificial barriers within psychology, such as between learning and the context in which learning occurs (Brown, Collins, & Duguid, 1989; Greeno, 1989) or between learning and development (Chi, 1978; Glaser, 1984).

Renewed linkages between the psychologies of learning and development are particularly noteworthy. Understanding how abilities develop is central to the task of understanding what abilities are. It is no accident that qualitative advances in our understanding of the mental processes which produce intelligent performances have come from those who studied the development of intelligence rather than those who focused exclusively or primarily on the organization of individual differences at a particular point in time. Much of this can be explained by a closer examination of the type of task typically studied by the developmentalist.

All scientific measurements of intelligence that we have at present are measures of some product produced by the person or animal in question, or of the way in which some product is produced [italics added]. A is rated as more intelligent than B because he produces a better product, essay written, answer found, choice made, completion supplied or the like, or produces an equally good product in a better way, more quickly or by inference rather than by rote memory, or by more ingenious use of the material at hand [italics added]. (E. L. Thorndike et al., 1926, p. 11-12)

Thorndike et al. (1926) here describe two types of tasks: tasks which permit inferences about the nature of intelligence from the type of response made (often a
qualitative judgment) and tasks which permit inferences about the rank order of individuals in ability by counting up the number of responses scored "correct" (usually a quantitative judgment). Psychometrics has understandably followed the quantitative route. Items are scaled for difficulty and examinees are ranked by how far up the ladder they can climb. Developmentalists from Piaget to Siegler have followed the qualitative path. The same problem is presented to all children and their developmental level is inferred from the sophistication of the response given. Indeed, early efforts to develop tests which provided a qualitative assessment of intelligence, such as the tests of Healy and Fernald (1911) or even the Binet scale of 1905, "did not emphasize the objective score which the child made so much as his general behavior and the way in which he went about the tasks which were set him" (Freeman, 1926, p. 108). However, judgments about process were clearly less dependable than judgments about whether the subject gave a keyed response, and so qualitative assessments of process were quickly displaced by quantitative assessments of product. Furthermore, tests which provided a score that could be immediately ranked better fit the requirements of a burgeoning test industry that was more interested in identifying who was intelligent than in understanding what intelligence was.

By the 1970s, however, cognitive psychologists had developed new methods for testing inferences about process—methods that were more sophisticated and objective than clinical judgments. Many tried to apply these new methods for detecting process to experimental tasks modeled after existing intelligence tests. Of all the "strange ironies" which have attended the history of mental testing (Cronbach, 1975), none is stranger than the attempt to apply powerful methods for detecting individual differences in processing strategy to a class of test-like tasks carefully pruned of such differences. It is a tribute to the power of the methods and the ingenuity of the researchers who employed them that anything interesting was found at all. Perhaps process analyses would be more successful in revealing interesting individual differences in process if they were to be applied to tasks deliberately designed to elicit such differences than to tasks modeled after existing mental tests.

Summary and Evaluation

Summaries broader than the scope of this paper are available (see Snow & Lohman, 1989; Sternberg, 1985), but several themes emerge in all of them.

First, much of the optimism about the potential impact of cognitive psychology on the study of human intelligence (e.g., Hunt et al., 1973; Sternberg, 1977) has been tempered by experience. Hunt now sees some fundamental incompatibilities between the correlational and experimental camps in psychology. He notes:

Cronbach [1957] thought that general theories of psychological process ought not to ignore individual differences, and vice versa. He was right, and in a general sense the union of the camps is well underway. In my opinion...the way to achieve the scientific union is to concentrate on understanding how individual differences variables, such as age, sex, genetic constitution, and education, influence the processes of cognition. It does not seem particularly fruitful to try to derive the dimensions of... [a trait model] of abilities from an underlying process theory. (Hunt, 1987, p. 36)
Like Hunt, Sternberg has also modified his views, although he sees more compatibility than Hunt. In 1977, Sternberg described a method for testing information-processing models of tasks that he called componential analysis. He then compared his method of componential analysis with factor-analytic methods for understanding abilities and found the latter seriously wanting. More recently, he has claimed that “cognitive approaches to intelligence are basically compatible with psychometric and other approaches” (1985, p. 108), each better suited to addressing different questions about the same phenomenon. Sternberg (1985) argues that his triarchic theory recognizes the contributions not only of the correlational and the information-processing approaches to the study of intelligence but also of theorists such as Berry (1972) and E. L. Thorndike et al. (1926) who point out that the list of behaviors and accomplishments valued as “intelligent” varies over cultures and contexts.

The conclusion that trait and process approaches are in some ways fundamentally incompatible may seem overly pessimistic. Nevertheless, it at least acknowledges that the two approaches make completely different demands on the basic person by item data matrix. Each partitions the data matrix in completely different ways. The trait theorist focuses on variation in row means whereas the experimentalist focuses on variation in column means. The trait theorist is concerned with covariances computed over persons whereas the experimentalist should be more concerned with covariances computed over items. It is possible—even likely—to propose a processing model that does an excellent job of accounting for variability in item difficulties or latencies, either for all subjects or separately for each subject, and yet have no explanation for individual differences on the task. On the other hand, the trait theorist constructs measures of broad abilities by making items (or subtests) as heterogeneous as possible (Spearman, 1927; Humphreys, 1985), thereby making a process analysis of the test either impossible or so general that it is uninformative. Thus, the two approaches are in some ways complementary but in other ways incompatible (Ippel & Lohman, 1990).

There has been a similar tempering of enthusiasm about the prospects for an easy victory over the problem of human intelligence in other quarters of cognitive science—particularly AI. Increasingly, those who have attempted to develop artificially intelligent systems have come to question their efforts and the constraints that the digital computer has placed on their work. In a summary of this recent history of AI, Dehn and Schank (1982) note, “Arrogance about the potential superiority of machine-specific intelligence slowly gave way to a growing respect for human intelligence and its operation. Characteristics of human intelligence...that had at first seemed to be weaknesses began to be recognized as strengths” (p. 354). For example, humans tend not to consider all aspects of a problem or to generate and evaluate all possible answers to a problem before deciding upon a course of action. Computers are easily programmed with algorithms that painstakingly consider all factors in a problem before choosing the best answer. However, the computer begins to drown in computation as problems increase in complexity, such as when the input is a visual scene or when the number of alternatives that could be generated is unlimited, as in a chess game. Further, this problem will not be solved by building computers with greater computational speed and power. Therefore, AI has shifted from programs that solve problems by brute force to programs modeled after the “satisficing” sort of rules of thumb.
humans use—balancing effort and time against expected payoff—in complex situations.

The recent shift to parallel-processing computers and to models of cognition that conform to current theories of brain function takes an even larger step away from the conventional digital computer and the constraints it imposes on efforts to model human cognition. However, some predict that even these efforts are doomed to fail, either because human cognition is not rule bound (Dreyfus & Dreyfus, 1986) or because higher level cognitive processes such as judgment and reasoning can be influenced by one’s beliefs, values, and intentions (Pylyshyn, 1984; Fodor, 1981).

In short, there has been a growing respect for human intelligence, and a realization that it will not yield to ready explanation by the methods of cognitive science any more than it yielded to ready explanation by the method of factor analysis. Yet factor analysis contributed—and continues to contribute (Carroll, in press: Gustafsson, 1984)—to our understanding of human intelligence. Cognitive science will also continue to contribute to our understanding in spite of the dire warnings of the pessimists and in spite of difficulties already encountered. But it will do so with a little less arrogance and, hopefully, with a little greater appreciation for the contributions of Binet, E. L. Thorndike, and others who have traveled this path before.

Notes

1 Special abilities often improve the prediction when samples are large or restricted on general ability (R. L. Thorndike, 1986). Note, too, that the issue is not general versus special abilities but whether to give each ability factor a unique weight or to give all the same weight in forming a single composite to be correlated with a criterion. It has long been known that a weighted average differs little from a simple average (Burt, 1907, cited in Butcher, 1968, p. 68). Instability of regression weights for correlated predictors demands it. Pooling correlations from different studies (e.g., Hunter, 1986) further exaggerates the role of general abilities (Linn, 1986). Finally, multiple aptitude batteries can still provide important information for guidance (Tyler, 1986).

2 Like many cognitive psychologists, Anderson (1985) usually uses the word intelligence as a synonym for cognition, not the individual difference construct associated with intelligence tests. The implications of this view for an individual difference interpretation are outlined in the third section of this paper.

3 For example, Freeman (1926) notes the need “to identify the mental processes which are measured by [existing ability tests]” (p. 127). He also provides a remarkably balanced summary of early research on intelligence.

4 Norman (1986) claims that the architecture of the digital computer was heavily influenced by the designers’ tacit theories of human cognition. Nevertheless, many who came later turned the metaphor around and looked for parallels between physical structures in the computer and psychological structures.

5 There are several intermediate cases as well. For example, Cronbach (1977) argues that “intelligence” is an abstraction much like “efficiency”. On this view, one cannot locate production efficiency in a particular department of a factory; rather, it is a term that describes the overall functioning of the system relative to comparable factories. Another possibility is that intelligence is something like Spearman’s (1927) mental energy or Jensen’s (1982) neural efficiency. Once again, one could not isolate “intelligence” in particular processes, but one might equate it with some general characteristics of cognition, such as attentional resources or speed of processing.
Fancher (1985) offers a fascinating historical perspective on the controversy. Using biographical sources, he traces the conflict from the disparate life experiences of John Stuart Mill and Frances Galton, through the lives of the major players in this controversy, to the recent debates between Kamin and Eysenck.

Humphreys (1986) aptly describes those who openly espouse environmental explanations for intelligence but who then assume some biological capacity not measured by existing intelligence tests that would be assessed by a properly constructed test as "closet hereditarians." The description seems also to apply to some cognitive scientists.

Carroll (1980) suggests that the correlation with verbal ability may be more parsimoniously attributed to a general or perceptual speed dimension. In a hierarchical model, however, factors such as perceptual (or clerical) speed, memory span, and fluency are located below verbal comprehension and thus represent specific verbal abilities. Carroll's critique is troublesome only if one views verbal comprehension as the sole verbal ability (see Snow & Lohman, 1989).

Low correlations with external criteria for all component scores except the intercept parameter is a statistical necessity unless task scores have poor internal consistency. This point is discussed below and in greater detail in Ippel and Lohman (in preparation). Thus, low correlations between components and other variables do not invalidate the models, although they do challenge the goal of estimating component scores for individuals.

For the there-is-nothing-new-under-the-sun folks, E. L. Thorndike et al. (1926) proposed that the various "products" of the human intellect be more systematically sampled from tests that differed in content ("including situations containing other human beings," p. 20) that required different "internal... processes" or "operations performed with the words, numbers, pictures, and other content" (p. 21).

Although procedural knowledge is said to be developed out of declarative knowledge, Anderson uses the term "procedural knowledge" more restrictively than some theorists. Knowledge of how to do something that is not yet compiled (or automatized) would be called declarative knowledge. Clearly, one can have declarative knowledge of a procedure or can have proceduralized that knowledge and not have a declarative representation of it, or one could have both.

As previously suggested, more informative process analyses demand tasks that allow ready inference of how subjects solved a problem, or what knowledge they brought to bear on it by the type of response they gave, not by the presence or absence of a correct response. In other words, the fundamental problem should be one of response categorization, not response scoring. Analyses of individual differences in response latencies introduce even more problems, such as what to do with error-response latencies or how to equate subjects on speed-accuracy trade-off. These problems are routinely ignored or incorrectly dismissed (for further discussion, see Lohman, 1989).

References


David F. Lohman


368


and their correlates. In M. P. Friedman, J. P. Das, & N. O’Connor (Eds.), Intelligence and learning (pp. 345–362). New York: Plenum Press.


Author

DAVID F. LOHMAN is Associate Professor, Division of Psychological and Quantitative Foundations, College of Education, The University of Iowa, Iowa City, IA 52242. He specializes in human abilities, cognitive psychology, and educational measurement.