review vs conceptual analysis
hard vs soft core
IP = "not European"

① Add Brian
language
connectionists

② Trace advances
over 15 yrs.

DEVELOPMENTAL PSYCHOLOGY
AN ADVANCED TEXTBOOK
Third Edition

Edited by
Marc H. Bornstein
Michael E. Lamb
National Institute of Child Health
and Human Development

LEA
1992
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Hillsdale, New Jersey
Hove and London

5 Information-Processing Approaches to Cognitive Development

David Klahr
Carnegie Mellon University

INTRODUCTION

Few psychologists would disagree with the claim that cognition involves the processing of information or that cognitive development involves changes in the content, structure, and processing of information. Indeed, since the 1970s, most of what has been discovered about children's thinking deals, in one way or another, with how they process information. However, if you asked different developmental psychologists to identify examples of information-processing approaches to cognitive development, you would probably find some interesting and important differences in their responses. Some might include the Piagetian and neo-Piagetian research that focuses on structures and structural changes. Others would limit the information-processing label to research that uses computer simulation to model developmental phenomena. Still others might point to the distinction between "classic" symbol-oriented information-processing theories (Newell & Simon, 1972) and more recent connectionist approaches (Bechtel & Abrahamsen 1991; Rumelhart & McClelland, 1986) to computational modeling of cognitive changes, arguing that only the latter are really suitable for modeling developmental processes (McClelland, 1989).

The nearly universal acceptance of the term *information processing,* when combined with diverse interpretations of its meaning, can conspire to bewilder and perplex the student of cognitive development. Which research paradigms exemplify information processing approaches? What are their merits? What have we learned about cognitive development from them? The goals of this chapter are twofold: (a) to describe the characteristic features of information-processing approaches to cognitive development, and (b) to illustrate what we have learned about children's cognitive development from the research characterized by different combinations of these features.

Defining "Information-Processing Approaches" to Cognitive Development

In psychology, as in all other scientific fields, research is carried out within the constraints of a set of basic theoretical assumptions and according to widely accepted methodological practices. Because they determine what questions will be asked and how they will be answered, these assumptions and practices can profoundly affect our understanding of cognition and its development. Information-processing approaches to cognitive development have their own characteristic set of theoretical assumptions and methodological practices. In this chapter, I attempt to reduce these to a manageable few: They are listed in Table 5.1. I organized this chapter around the list of assumptions and practices presented in Table 5.1. For each of the entries listed, I describe several studies, models, or findings that exemplify that entry. Although any particular example represents a combination of features listed in Table 5.1, I have attempted to locate them where they best illustrate the item being discussed.

The assumptions (A1, A2, and A3) and the practices associated with them (P1–P4) vary along what I have called a *soft-core* to *hard-core* continuum (Klahr, 1989). For example, some of the studies described in this chapter, while accepting assumptions A1 to A3, are not at all specific about any of

TABLE 5.1
Assumptions and Practices of Information-Processing Approaches to the Study of Cognitive Development

Theoretical Assumptions

A1: Children's mental activity involves processes that manipulate symbols and symbol structures.

A2: These symbolic processes operate within an information-processing system having identifiable properties, constraints, and consequences.

A3: Cognitive development occurs via self-modification of the information-processing system.

Methodological Practices

P1: Use of highly detailed analyses of the environment facing the child on specific tasks.

P2: Use of formal notational schemes for expressing complex, dynamic systems.

P3: Measuring the time-course of cognitive processing over both relatively short durations (chronometric analysis) and medium durations (microgenetic studies).

P4: Use of high-density data from error patterns and protocols to induce and test complex models.

them, while other studies include computer-simulation models of how children accomplish some tasks. The former would be soft-core examples, and the latter would be hard core. With respect to the methodological practices, the soft end of P2 would involve the use of flow-charts and diagrams to describe a model of children's thinking, whereas the hard end would involve a computational model. In addition to varying along the hardness continuum, different examples vary with respect to how many of the assumptions and practices they reflect. Rather than treating these descriptors as a set of defining properties, they should be interpreted as likely properties of typical examples. That is, the information-processing notion itself is better expressed in terms of family resemblance than as an idea having clear defining properties. Before visiting the members of this family, it may be useful to explore a bit of its genealogy.

Origins

In the early 1970s, Roger Brown reviewed the previous two decades in an attempt to identify the forces that revitalized research in cognitive development in the late 1950s. One of them was the creation of computer simulation models of cognitive processes in adults. As Brown (1970) said:

> Since machines—hardware—could accomplish information processing of great complexity, it was obviously perfectly scientific and objective to attribute such processing to the human brain. Why limit the mind to association by contiguity and reinforcement when the computer, admittedly a lesser mechanism, could do so much more? Computers freed psychologists to invent mental processes as complex as they liked. (pp. ix–x)

The other force identified by Brown was America's discovery of Jean Piaget:

> computer simulation, psycholinguistics, curriculum reform, and mathematical models altered our notions of the scientific enterprise in such a way to cause us to see Piaget as a very modern psychologist. To see that he was, in fact, the great psychologist of cognitive development. (p. x)

Ten years prior to Brown's acknowledgment of the relevance and impact of computational models to the topics first addressed by Piaget, Herbert Simon (1962) had suggested the general form of an information-processing approach to cognitive development:

> If we can construct an information-processing system with rules of behavior that lead it to behave like the dynamic system we are trying to describe, then this system is a theory of the child at one stage of the development. Having

> described a particular stage by a program, we would then face the task of discovering what additional information-processing mechanisms are needed to simulate developmental change—the transition from one stage to the next. That is, we would need to discover how the system could modify its own structure. Thus, the theory would have two parts—a program to describe performance at a particular stage and a learning program governing the transitions from stage to stage. (pp. 154-155)

Simon's suggestion contained two ideas that departed radically from the then prevailing views in developmental psychology.[1] The first idea was that theories about thinking could be stated as computer programs. These "computational models of thought," as they came to be called, have one important property that distinguishes them from all other types of theoretical statements: They independently execute the mental processes they represent. That is, rather than leaving it to the reader to interpret a verbal statement about what is involved in an analogical mapping or a memory search or a match between two symbols, computational models actually do the mapping, searching, and matching so that the complex implications of multiple processes can be unambiguously derived. The second idea in Simon's suggestion followed from the first: If different states of cognitive development could be described as programs, then the developmental process itself could also be described as a program that took the earlier program and transformed it into the later one. Such a program would have the capacity to alter and extend its own processes and structures. That is, it would be a computational model possessing some of the same self-modification capacities as the child's developing mind.

Today, the soft-core versions of Simon's two ideas form the cornerstone of a very large proportion of the research on both adult cognition and cognitive development. There is no question that in the field of adult cognition, information-processing approaches have had an enormous impact on both theory and methodology (Lachman, Lachman, & Butterfield, 1979; Palmer & Kimchi, 1986). One reason that the idea of thinking as information processing is so widespread is that it is highly nonspecific. As we shall see, different investigators draw quite different implications from this general notion. However, the hard-core version of A1 to A3—the "theory is the program" view expressed by Simon—has yet to become the dominant view in developmental psychology. Even among the many developmentalists who accept assumptions A1 to A3, there are relatively few who can point to examples in their own work of hard-core implementations of information-processing theories. Nevertheless, there is a general trend in the field toward "hardening the core," and throughout this chapter I refer to examples of the trend.

---

[1] It is easy to forget just where the field was in the early 1960s. Note that the Kendlers' famous work on reversal shift, which challenged the prevailing notions of S-R learning by suggesting that an internal mediator played a role in concept acquisition, was published in the same year as Simon's (Kendler & Kendler, 1962).

CHILDREN'S MENTAL ACTIVITY INVOLVES PROCESSES THAT MANIPULATE SYMBOLS AND SYMBOL STRUCTURES

The first of the assumptions listed in Table 5.1 is the most pervasive and, correspondingly, the most diffusely defined. In this section I illustrate the different meanings that assumption A1 has taken in the developmental literature. The examples will also describe some important findings about children's information processing. This is the pattern followed throughout the rest of the chapter: Specific examples illustrate the main features listed in Table 5.1, and describe a bit of what we know about how children think.

Symbolization in the most diffuse sense addresses the power of the child's representational capacity, without any concern with, or commitment to, how that capacity might be supported in a physical system. Examples include Piagetian accounts of "symbolic play" or imagery (Piaget, 1951). As information-processing accounts move along the soft to hard dimension, they use terms such as *symbol* and *symbol-structure* in ways that are both more mechanistic and more microscopic.

Newell (1980) defined the hard-core information-processing view of the role of symbols, symbol structures, and symbol manipulation in cognition. He defined a physical symbol system as one that:

> is capable of having and manipulating symbols, yet is also realizable within our physical universe . . . [This concept] has emerged from our growing experience and analysis of the computer and how to program it to perform intellectual and perceptual tasks. The notion of symbol that it defines is internal to this concept of a system. Thus, it is a hypothesis that these symbols are in fact the same symbols that we humans have and use everyday of our lives. Stated another way, the hypothesis is that humans are instances of physical symbol systems, and by virtue of this, mind enters into the physical universe. (p. 136)

The fundamental property of a symbol is that it can designate[2] something else (represented as a symbol structure). Such symbols comprise the elementary units in any representation of knowledge including sensory-motor knowledge or linguistic structures. Philosophical distinctions between dense and articulated symbols (Goodman, 1968) or personal and consensual symbols (Kolers & Smythe, 1984) emphasize the likelihood of idiosyncratic symbol structures for specific individuals, and the difference between internal symbol structures and their external referents. However, they are entirely consistent with Newell's Physical Symbol System hypothesis.

---

[2] As Newell noted, many terms similar to designate could be used here: refer, denote, name, stand for, mean, etc.

Are Preschoolers Presymbolic?

These terminological distinctions become important when one asks questions about the developmental course of symbolic capacity, because unless one is specific about the sense in which symbolic is intended, one will find inconsistent and contradictory results in the literature. A very clear distinction is made by DeLoache (1987) in her investigations of preschoolers' ability to use one thing to represent another.

DeLoache investigated this question by presenting children with a scale model of a full-sized room, and then determining the extent to which children understood the correspondences between the two. In one series of studies, children were familiarized with a room filled with assorted furniture, and they watched while a toy was hidden. Then they were shown a scale model of the room, and asked to find a miniature version of the toy. They were instructed that the miniature toy was in the "same place" in the model as the full-sized one was in the full-sized room, and they were instructed to try to find it. Following the retrieval from the model, the children were asked to find the original item. (This was done as a memory check to make sure that children had not forgotten where the toy was originally hidden.) For some children the role of the model and the full-size room was as just described, and for others it was reversed. Two age groups were used: a 2.5-year-old group and a 3-year-old group.

The results were dramatic. The older group found the toy on about 80% of the trials, while the younger group found it on less than 20%. Both groups could remember the original hiding place on about 80% of the trials, so faulty memory cannot explain the results. Nor was there any effect for whether the room or the model was used for the original hiding (with the model or room, respectively used as the retrieval location). In a second experiment, DeLoache used a photograph of the room, instead of the actual room, to indicate where the object was hidden. With this change, 2.5-year-old children were able to perform at nearly the same level as the 3-year-olds in the first experiment. DeLoache notes that this outcome is "directly contrary to the standard view of the efficacy of pictures versus real objects" (p. 1557).

Taken as a whole, DeLoache's results demonstrate an abrupt improvement between 30 and 36 months in children's ability to understand the symbolic relations between a model of a room and the real room. DeLoache summarizes this as a milestone in "the realization that an object can be understood both as a thing itself and as a symbol of something else" (DeLoache, 1987, p. 1556), and she notes that the younger children fail "to think about a symbolic object both as an object and as a symbol" (p. 1557). Thus, at the global (or conventional) level, DeLoache's results suggest that the 2.5-year-old children are presymbolic (at least on this task.) But it is clear that if one were to formulate detailed models of both the younger and older children's knowledge about this task, one would, in both cases, postulate systems that had the ability to process symbols at the microscopic level defined above. Thus, even in an ingenious research program—such as DeLoache's—directed at discovering rapid changes in the "symbolic functioning of very young children," the assumption of underlying symbol-processing capacity remains.

Knowledge Structures as Symbol Structures

Assumptions about the centrality of symbol structures are implicit in the "knowledge is power" approach to cognitive development. The general goal of this line of work is to demonstrate that much of the advantage adults have over children derives from their more extensive knowledge base in specific domains, rather than from more powerful general processes. But as Chi and Ceci (1987) comment, "saying that young children have less knowledge than older children or adults borders on triviality. The sheer quantity of knowledge, although important, is not nearly as important as how that knowledge is structured" (p. 115). Chi's studies (1976, 1977, 1978) provide convincing evidence for the influence of both more and better structured knowledge. In all of these investigations, Chi found that children who have more richly connected, domain-specific, knowledge than adults (e.g., children who have more knowledge than their adult counterparts about chess or dinosaurs or classmates' faces) outperform their adult counterparts on a range of tasks in which access to that knowledge is a determining factor in performance.

For example, in one study Chi (1978) examined the differences between adults and 10-year-olds on two tasks: a conventional digit span task and memory for chess positions. The children were all experienced chess players, and on a standard chess task, they performed slightly better than did the adults, who were novice chess players. The digit span task, presented to both groups, yielded the standard result: Adults' spans are greater than children's. The criterion chess task was memory for the location of pieces from various mid-game positions. Here, the children outperformed the adults. This general effect has been replicated many times, with a wide range of materials (Barrett, 1978; Lindberg, 1980).

In all of these, and related, studies, the major explanatory variable is access to symbolic structures (chunks, semantic nets, etc.) that support the superior performance of the children. Although, as Chi and Ceci (1987) argue, there exists "a lack of consensus on precisely what *structure* means" (p. 129), just about every definition includes the ability for one part of the structure to be connected to (or provide access to, or evoke, or derive inferences or generalizations from) some other part of the structure. And underlying any of these interpretations are just the kind of symbol structures described in Newell's concept of physical symbol systems.

SYMBOLIC PROCESSES OPERATE WITHIN AN INFORMATION-PROCESSING SYSTEM HAVING IDENTIFIABLE PROPERTIES, CONSTRAINTS, AND CONSEQUENCES

Developmentalists interested in a variety of cognitive processes have generally adopted the view of the adult information-processing system that emerged in the late 1960s and early 1970s (Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972; Norman, Rumelhart, & the LNR Research Group, 1975). However, as in all other aspects of information processing, there is a wide range of interpretations and applications of the general idea of information processing that vary along the hard to soft dimension. In this section, I begin with a general description of a widely accepted view of the human information-processing system. Then I go on to describe some specific hard-core computational models. Following that I discuss some soft-core examples of information-processing systems.

The Organization of the Human Information-Processing System

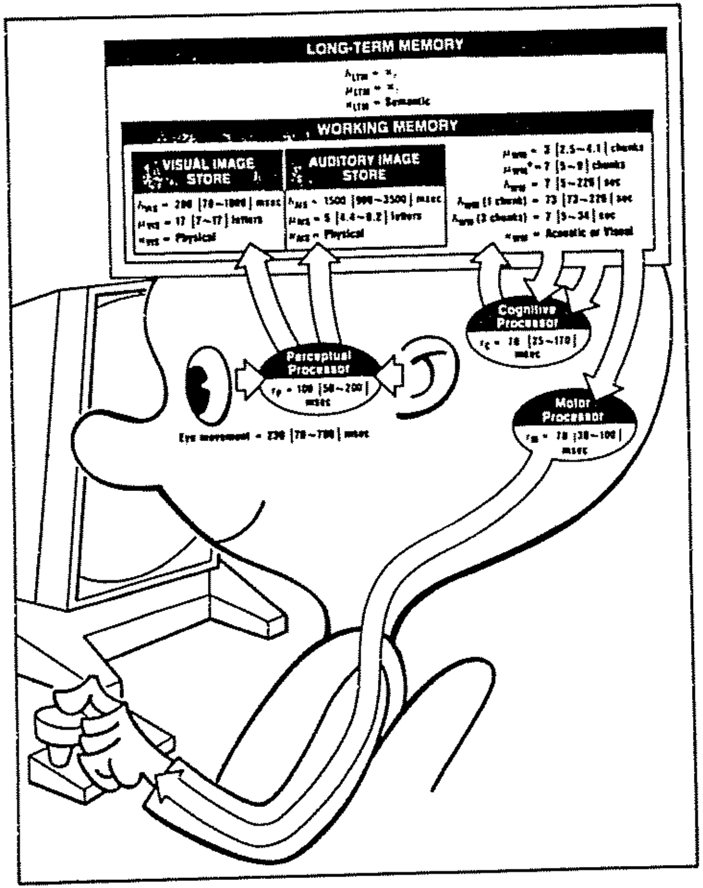
The standard description of the human information-processing system includes several sensory buffers (e.g., "iconic" memory, an "acoustic buffer"), a limited capacity short-term memory (STM), and an unlimited, content-addressable long-term memory). This characterization is unabashedly derived from, and analogous to, the gross functional features of computer architectures. Nevertheless, as I argue later in this chapter, this does not imply that information-processing psychologists believe that the brain is structurally organized like a computer.

Newell (1972, 1973, 1981) originated the idea of a *cognitive architecture* of the mind. The idea has gone through successive elaborations, one of which is described in Card, Moran, and Newell's (1983) proposal for what they called the Model Human Processor (MHP). This is a model of the human information-processing system that includes not only the gross organization of the different information stores and their connections, but also estimates of processing rates and capacities. MHP was designed to facilitate predictions about human behavior in a variety of situations involving interactions between humans and computers. It was based on a vast amount of empirical data on human performance in perceptual, auditory, motor, and simple cognitive tasks.

Their model is illustrated in Fig. 5.1a and 5.1b. It includes a long-term memory, a working memory, two perceptual stores for visual and auditory information, and three subsystems for cognitive, motor, and perceptual processing. For each of these stores, there are associated estimates of storage capacity, decay times, cycle times, and the type of code as well as connectivity to the rest of the system.

> The perceptual system consists of sensors and associated buffer memories, the most important buffer memories being a Visual Image Store and an Auditory Image Store to hold the output of the sensory system while it is being symbolically coded. The cognitive system receives symbolically coded information from the sensory image stores in its Working Memory and uses previously stored information in Long-Term Memory to make decisions about how to respond. The motor system carries out the response. As an approximation, the information processing of the human will be described as if there were a separate processor for each subsystem: a Perceptual Processor, a Cognitive Processor, and a Motor Processor. For some tasks (pressing a key in response to a light) the human must behave as a serial processor. For other tasks (typing, reading, simultaneous translation) integrated, parallel operation of the three subsystems is possible, in the manner of three pipelined processors: information flows continuously from input to output with a characteristically short time lag showing that all three processors are working simultaneously.
>
> The memories and processors are described by a few parameters. The most important parameters of a memory are μ, the storage capacity in items, δ, the decay time of an item, and κ, the main code type (physical, acoustic, visual, semantic). The most important parameter of a processor is τ, the cycle time. (Card et al., 1983, pp. 24-25)

Although the MHP was formulated to account for the perceptual and motor behavior of adults interacting with computers, it is a good example of the more general attempt to formulate a cognitive architecture of the mind. More specifically, it is a very successful integration of a general information-processing orientation with the constraints provided by a massive amount of experimental data on human performance. To date, no one has proposed a "kiddie" version of MHP, although some attempts have


LONG-TERM MEMORY
$\delta_{LTM} = \infty$
$\mu_{LTM} = \infty$
κ_{LTM} = Semantic
WORKING MEMORY
VISUAL IMAGE STORE
δ_{VIS} = 200 [70~1000] msec
μ_{VIS} = 17 [7~17] letters
κ_{VIS} = Physical
AUDITORY IMAGE STORE
δ_{AIS} = 1500 [900~3500] msec
μ_{AIS} = 5 [4.4~6.2] letters
κ_{AIS} = Physical
μ_{WM} = 3 [2.5~4.1] chunks
μ_{WM}^* = 7 [5~9] chunks
δ_{WM} = 7 [5~226] sec
δ_{WM} (1 chunk) = 73 [73~226] sec
δ_{WM} (3 chunks) = 7 [5~34] sec
κ_{WM} = Acoustic or Visual
Perceptual Processor
τ_P = 100 [50~200] msec
Cognitive Processor
τ_C = 70 [25~170] msec
Motor Processor
τ_M = 70 [30~100] msec
Eye movement = 230 [70~700] msec

(a)

P0. *Recognize-Act Cycle of the Cognitive Processor.* On each cycle of the Cognitive Processor, the contents of Working Memory initiate actions associatively linked to them in Long-Term Memory; these actions in turn modify the contents of Working Memory.

P1. *Variable Perceptual Processor Rate Principle.* The Perceptual Processor cycle time τ_P varies inversely with stimulus intensity.

P2. *Encoding Specificity Principle.* Specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored.

P3. *Discrimination Principle.* The difficulty of memory retrieval is determined by the candidates that exist in the memory, relative to the retrieval clues.

P4. *Variable Cognitive Processor Rate Principle.* The Cognitive Processor cycle time τ_C is shorter when greater effort is induced by increased task demands or information loads; it also diminishes with practice.

P5. *Fitt's Law.* The time T_{pos} to move the hand to a target of size S which lies a distance D away is given by:

$$T_{pos} = I_M \log_2 (D/S + .5),$$

where I_M = 100 [70 ~ 120] msec/bit.

P6. *Power Law of Practice.* The time T_n to perform a task on the nth trial follows a power law:

$$T_n = T_1 n^{-\alpha},$$

where α = .4 [.2 ~ .6].

P7. *Uncertainty Principle.* Decision time T increases with uncertainty about the judgement or decision to be made:

$$T = I_C H,$$

where H is the information-theoretic entropy of the decision and I_C = 150 [0 ~ 157] msec/bit. For n equally probable alternatives (called Hick's Law),

$$H = \log_2 (n + 1).$$

For n alternatives with different probabilities, p_i, of occurrence,

$$H = \sum_i p_i \log_2 (1/p_i + 1).$$

P8. *Rationality Principle.* A person acts so as to attain his goals through rational action, given the structure of the task and his inputs of information and bounded by limitations on his knowledge and processing ability:

Goals + Task + Operators + Inputs + Knowledge + Process-limits → Behavior

P9. *Problem Space Principle.* The rational activity in which people engage to solve a problem can be described in terms of (1) a set of states of knowledge, (2) operators for changing one state into another, (3) constraints on applying operators, and (4) control knowledge for deciding which operator to apply next.

(b)

FIG. 5.1. (a) The Model Human Processor—memories and processors. Sensory information flows into Working Memory through the Perceptual Processor. Working memory consists of activated chunks in Long-Term Memory. The basic principle of operation of the Model Human Processor is the *Recognize-Act Cycle of the Cognitive Processor* (PO in Fig. 5.1b). The Motor Processor is set in motion through activation of chunks in Working Memory (from Card, Moran, & Newell, 1983). (b) The Model Human Processor—principles of operation.

been made to chart the developmental course of some of its parameters. (See the description of Kail's work in a later section.)

Production Systems

Cognition has both a serial and a parallel aspect to it. At both the underlying neural level, where massively parallel computations are occurring, and at the perceptual–cognitive interface, where the sense organs encode the external world for further processing by higher order mental processes, there must be a high degree of parallelism. On the other hand, both rational thought and motor acts from speech to locomotion require a certain degree of seriality. These considerations led Newell and Simon (1972) to propose a formalization of high-order mental processes in terms of condition-action rules called productions. Newell (1973) implemented this idea as a programming language, called PSG,[3] for creating computational models as running production systems.

Production systems are a class of computer-simulation models stated in terms of condition–action rules. A production system consists of two interacting data structures: (a) A *working memory* consisting of a collection of symbol structures called working memory *elements;* (b) A *production memory* consisting of condition–action rules called *productions,* the conditions of which describe configurations of working memory elements and the actions of which specify modifications to the contents of working memory. Production memory and working memory are related through the *recognize–act* cycle, which is comprised of three distinct processes:

1. The *match* process finds productions the conditions of which match against the current state of working memory. The same rule may match against working memory in different ways, and each such mapping is called an *instantiation.* When a particular production is instantiated, we say that its conditions have been satisfied. In addition to the possibility of a single production being satisfied by several distinct instantiations, several different productions may be satisfied at once. Both of these situations lead to conflict.
2. The *conflict resolution* process selects one or more of the instantiated productions for applications.

---

[3] PSG is an acronym for "Production System, version G." Although this was the first publically distributed general-purpose system for running production systems on computers, the "version G" implies that Newell had deemed its six precursor versions unsuitable for public consumption. For a brief account of the genealogy of production system languages see Neches, Langley, and Klahr (1987).

3. The *act* process applies the instantiated actions of the selected rules. Actions can include the modification of the contents of working memory, as well as external perceptual-motor acts.

Production systems can be thought of as complex, dynamic stimulus-response pairs in which both the S and the R involve symbolic structures. They provide both a parallel associative recognition memory, on the condition side, and a serial response on the action side. The basic recognize-act process operates in cycles, with one or more rules being selected and applied, the new contents of working memory leading another set of rules to be applied, and so forth.

Both Newell's PSG, and Anderson's (1983) ACT*, which combined production systems with semantic nets, provided computational languages for formulating cognitive models. More important, these systems were theoretical extensions of the standard model into some very specific proposals about how the human cognitive architecture is structured. These proposals took the form of a type of computational architecture—the production system—and production systems have since been used to model several aspects of cognitive development.

Production-System Models of Children's Performance

For developmentalists, one of the most valuable features of production systems is their potential to model the change process itself: their potential for self-modification. Later in this chapter, I explain why self-modification is such an important and powerful feature of hard-core information-processing models, and I describe some approaches that exploit this self-modification capacity. But first, I describe production-system models of children's performance at specific levels of development. These models, even though cast only as models of different performance levels, rather than as models of transition processes, serve useful functions.

In this section I describe four different ways in which *non* self-modifying production systems have been used to model children's performance. The first example illustrates how production systems can be matched to chronometric data to produce some estimates of the duration of elementary components of the recognize-act cycle. The second example illustrates one of the most valuable features of production systems for modeling cognitive development: the ease with which different performance levels can be represented by a family of models having different production sets. The third example focuses on how production systems can include encoding and performance productions in the same general format, and the final example illustrates a kind of vertical integration in a production-system model that represents several levels of knowledge from general principles down to specific encoding rules.

Quantification: Matching Production Firings to Chronometric Data. Production-system models of thinking were initially developed to account for the verbal protocols generated by subjects working on puzzles requiring several minutes to solve (Newell, 1966). However, a much finer temporal grain of analysis was used in the first production-system models that actually ran as computer simulations. Newell (1973) introduced PSG in the context of the Sternberg memory-scanning paradigm (described later in this chapter). The same volume (Chase, 1973) included a description of a model, written in PSG, of elementary processes for quantification: subitizing, counting, and adding (Klahr, 1973). Both of these models were atypical of most subsequent production-system models in that they attempted to account for chronometric data in terms of the dynamic properties of the production-system execution cycle. That is, they estimated the duration of specific microprocesses within the recognize–act cycle (such as the time to do a match, or the time to execute an action) by relating the number of such microprocess executions to the reaction-time data.

Although neither of these early models dealt with developmental data, the model of elementary quantification processes was subsequently elaborated into one that did deal with the differences in subitizing rates between children and adults (Klahr & Wallace, 1976, Chapter 3 & 8). The elaboration included two distinct working memories: one corresponding to the traditional short-term memory, and the other corresponding to an iconic store. Accordingly, the condition elements in productions could refer to either of these information sources, and the time parameters associated with matches in the two stores differed.

By attempting to constrain the model-building process with the chronometric data from very different domains, both Newell's model and Klahr and Wallace's model converged on a gross estimate of the time duration for the basic production-system cycle time of between 10 and 100 ms. While this may seem to be a fairly loose parameter estimate, it is important to note that it is not 1 ms, nor is it 1000 ms. That is, if the production cycle is constrained, even within these broad limits, then one can evaluate the plausibility of particular production systems in terms of whether they exhibit—within an order of magnitude—the same absolute as well as relative temporal patterns as do the humans they are modeling.

Production Systems for Different Levels of Performance. Another use of production systems by developmentalists has been the sequence-of-models approach. The goal here is to produce a sequence of production-system models for a specific task such that each model represents a different level of performance. Once it has been demonstrated that the models can indeed produce the appropriate behavior at each level of performance, then one can examine the differences between successive models in order to infer what a transition mechanism would have to accomplish.

Baylor and Gascon (1974) did this type of analysis in their investigation of developmental differences in children's ability to do weight seriation. They presented children between the ages of 6 and 12 years old with a task in which the goal was to create an ordered series of identically appearing objects having different weights. Children could make pair-wise comparisons of the objects by using a balance scale, but they could not get an absolute measurement of an object's weight. Baylor and Gascon observed children's behavior as they weighed different pairs of objects and attempted to arrange them according to weight. From the sequence of children's pair-wise comparisons, Baylor and Gascon inferred a set of increasingly effective strategies. Each strategy was formulated as a production system having different collections of elementary components. Each of these production systems was implemented as a running computer program and the program's sequence of comparisons and the final outcome provided a good fit to individual children's sequences of object manipulations.

Klahr and Siegler (1978) used production systems in a different way: to take a soft-core information-processing model—one that had already shown an excellent fit to the children's performance—and extend it to a production-system format so as to get a better idea of its demands on short-term memory and its dynamic properties. Siegler had previously proposed an elegant analysis of rule sequences characterizing how children (from 3 years old to 17 years old) make predictions in several domains (Siegler, 1976, 1981), and the sequences were formulated as a series of increasingly elaborated binary decision trees. By recasting the rules as production systems, Klahr and Siegler were able to make a more precise characterization of what develops than was afforded by just the decision-tree representation. The following quotation from Klahr and Siegler (1978) conveys the level of detail that was facilitated by the production-system formulation.

> We can compare the four models [production system versions of Siegler's four "rule models"] at a finer level of analysis by looking at the implicit requirements for encoding and comparing the important qualities in the environment. Model I tests for sameness or difference in weight. Thus, it requires an encoding process that either directly encodes relative weight, or encodes an absolute amount of each and then inputs those representations into a comparison process. Whatever the form of the comparison process, it must be able to produce not only a same-or-different symbol, but if there is a difference, it must be able to keep track of which side is greater. Model II requires the additional capacity to make these decisions about distance as well

> as weight. This might constitute a completely separate encoding and comparison system for distance representations, or it might be the same system except for the interface with the environment.
>
> Model III needs no additional operators at this level. Thus, it differs from Model II only in the way it utilizes information that is already accessible to Model II. Model IV requires a much more powerful set of quantitative operators than any of the preceding models. In order to determine relative torque, it must first determine the absolute torque on each side of the scale, and this in turn requires exact numerical representation of weight and distance. In addition, the torque computation would require access to the necessary arithmetic production systems to actually do the sum of products calculations. (p. 80)

Representing the Immediate Task Context. One advantage of a production-system formulation is that it facilitates the extension of a basic model of the logical properties of a task to include the processing of verbal instructions, encoding of the stimulus, keeping track of where the child is in the overall task, and so on. For example, in their analysis of individual subject protocols on the balance scale, Klahr and Siegler proposed some models to account for some children's idiosyncratic—but consistent—response patterns. One of these models included not only the basic productions for a variant of one of Siegler's four models for balance scale predictions, but also a lot of other knowledge about the task context:

> The model represents, in addition to the child's knowledge about how the balance scale operates, her knowledge about the immediate experimental context in which she is functioning. The trial-by-trial cycle during the training phase comprises (1) observation of the static display, (2) prediction of the outcome, (3) observation of the outcome, (4) comparison of the outcome with the prediction, and (5) revision if necessary of the criterion. . . . This model utilizes, in one way or another, representation of knowledge about when and how to encode the environment, which side has *more* weight or distance, which side has a *big* weight or distance, what the current criterion value is, what the scale is expected to do, what the scale actually did, whether the prediction is yet to be made or has been made, and whether it is correct or incorrect. (Klahr & Siegler, 1978, p. 89)

This kind of model raises two issues that might otherwise escape notice. First, what kinds of knowledge are necessary to generate these different encodings? It has long been known that surface variations in tasks can cause wide variation in children's performance—even on the tasks purported to index developmental level, such as class inclusion (Klahr & Wallace, 1972). Production-system formulations avoid the arbitrary dichotomy between performance demands and the so-called logical properties of a task, and force an unambiguous specification of all the processing necessary to complete the task. Second, how much of the encoded knowledge (i.e., the contents of working memory) must be available at any one moment? That is, in order to do the task, how much working memory capacity is required? Case (1986) addresses this issue informally in his proposed procedures for quantifying tasks in terms of their demands on the Short-Term Storage Space (STSS). However, without a clear and principled specification of the grain-size and computational power of the routines that use the contents of STSS, it is difficult to apply his demand-estimating procedure to a new domain.

Multiple-Level Production Systems: From General Rules to Detailed Encodings. Klahr and Wallace (1976) describe a model of children's performance on Piaget's conservation of quantity task. Their model contains productions dealing with several different levels of knowledge. At the highest level are productions that represent general conservation rules, such as "If you know about an initial quantitative relation, and a transformation, then you know something about the resultant quantitative relation." (See Klahr & Wallace, 1973, for an elucidation of these conservation rules.) At the next level are productions representing pragmatic rules, such as "If you want to compare two quantities, and you don't know about any prior comparisons, then quantify each of them." At an even lower level are rules that determine which of several quantification processes will actually be used to encode the external display (e.g., subitizing, counting, or estimation). Finally, at the lowest level, are productions for carrying out the quantification process. These are the same productions that comprised the systems described earlier in our discussion about matching production systems to chronometric data.

Although I have described this system as if there were a hierarchy of productions, there is only the flat structure of a collection of productions. Each production simply checks for its conditions. If it fires, then it deposits its results in working memory. The hierarchy emerges from the specific condition elements in each production, which ensure that productions only fire when the current context is relevant.

Nontransition Models: A Summary. These four instances by no means exhaust the set of computer simulations of children's thinking processes. Rabinowitz, Grant, and Dingley (1987) summarize over a score of other computer simulation models relevant to cognitive development, including those that use non-production-system architectures, and including both state and transition models. The production-system models include work on seriation (Baylor, Gascon, Lemoyne, & Pother, 1973; Young, 1976) and subtraction (Young & O'shea, 1981). Computer simulations based on schema architectures have been proposed in the area of arithmetic (Greeno, Riley, & Gelman, 1984; Kintsch & Greeno, 1985; Riley, Greeno, & Heller, 1983) and language acquisition (Hill, 1983). Task-specific architectures have been used to model children's performance on addition (Ashcraft, 1987; Siegler, 1988), subtraction (Brown & VanLehn, 1982), and series completion (Klahr & Wallace, 1970a). As Rabinowitz and colleagues observe, only a handful of these models include self-modifying mechanisms. Nevertheless, the underlying assumption in all of the computer simulations is that by clarifying the nature of children's thought at any particular level of development, the requirements of a transition theory become better defined.

Other Computational Models[4]

Production systems are not the only kind of computational model used to model children's thinking. In some cases, the researcher is not conforming to any particular theoretical assumptions about cognitive architectures, but still has a theory that is sufficiently complex that only a computational model will enable him or her to derive predictions from it. In such cases the researcher simply chooses to focus on the main data structures and computational processes, and employs an atheoretical computational architecture in which to formulate and run the model.

Siegler and Shrager (1984) proposed such a model to account for an unusually rich data set based on 4- and 5-year-old children's performance on simple addition problems (with sums less than 10). The model, shown in Fig. 5.2, is based on two basic ideas: (a) Children will retrieve answers from memory to problems that they are very certain about, and they will use other strategies (such as counting on their fingers) when they are not so sure; (b) Each possible problem (m + n) has a distribution of possible responses associated with it (see Fig. 5.2a). Some problems (e.g., 1 + 2) have very sharply peaked distributions, so that a single answer (3), is strongly associated with the problem. Other problems have distributions of possible answers that are bimodal (e.g., both 5 and 7 are likely to be retrieved in response to the problem 3 + 4), or relatively flat (e.g., 5 + 3), so that several answers are weakly associated with the problem. Siegler and Shrager elaborated these rather general and intuitive notions into a computational model that both acquired the distributions of associations with

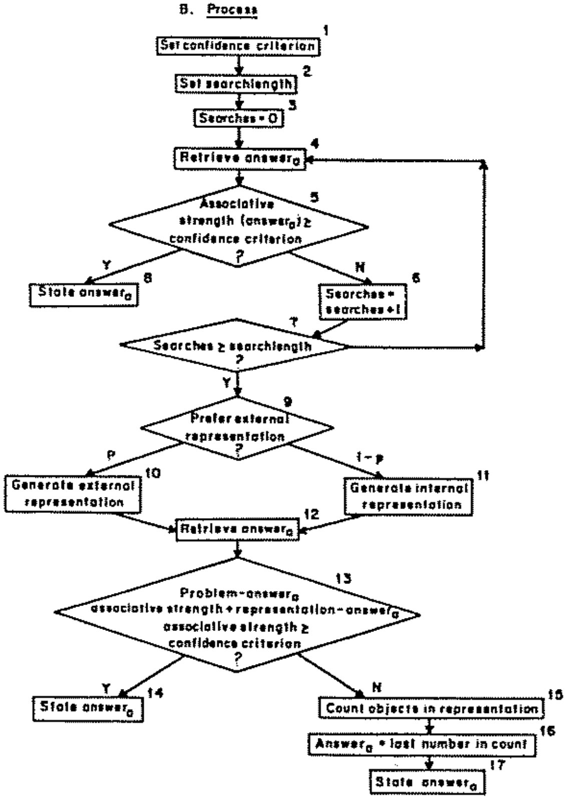
---

[4] The computational models described in this chapter are all variants of the symbol-oriented approach to cognition, in contrast to the connectionist (or parallel distributed processing—PDP—approach). In the penultimate section on "Constraints and Limitations," I discuss some of the potential contributions of connectionist approaches to information processing in children.

A. Representation (Associative Strengths)

PROBLEM	ANSWER												
	0	1	2	3	4	5	6	7	8	9	10	11	OTHER
1+1		.05	.86		.02		.02					.02	.04
1+2			.09	.70	.02		.04			.07	.02	.02	.05
1+3		.02		.11	.71	.05	.02	.02					.07
1+4					.11	.61	.09	.07				.02	.11
1+5					.13	.16	.50	.11		.02	.02		.05
2+1		.07	.05	.79	.05								.04
2+2	.02		.04	.05	.80	.04		.05					
2+3			.04	.07	.38	.34	.09	.02	.02	.02			.04
2+4		.02		.07	.02	.43	.29	.07	.07				.04
2+5		.02		.05	.02	.16	.43	.13			.02		.18
3+1		.02		.09	.79	.04		.04					.04
3+2			.09	.11	.11	.55	.07						.07
3+3	.04			.05	.21	.09	.48		.02	.02	.02		.07
3+4				.05	.11	.23	.14	.29	.02				.16
3+5				.07		.13	.23	.14	.18		.05		.20
4+1			.04	.02	.09	.68	.02	.02	.07				.07
4+2			.07	.09		.20	.36	.13	.07		.02		.07
4+3				.05	.18	.09	.09	.38	.09		.02		.11
4+4	.04			.02	.02	.29	.07	.07	.34		.04		.13
4+5					.04	.09	.16	.09	.11	.18	.11	.04	.20
5+1			.04		.04	.07	.71	.04	.04		.04		.04
5+2			.05	.20	.02	.18	.27	.25	.02		.02		
5+3			.02	.11	.09	.18	.05	.16	.23		.05		.11
5+4					.11	.21	.16	.05	.11	.16	.04		.16
5+5	.04					.07	.25	.11	.02	.04	.34	.04	.11

(a)


B. Process
Set confidence criterion
1
Set searchlength
2
Searches = 0
3
Retrieve answer$_a$
4
Associative strength (answer$_a$) ≥ confidence criterion ?
5
Y
N
State answer$_a$
8
Searches = searches + 1
6
Searches ≥ searchlength ?
7
Y
Prefer external representation ?
9
p
1 − p
Generate external representation
10
Generate internal representation
11
Retrieve answer$_a$
12
Problem-answer$_a$ associative strength + representation-answer$_a$ associative strength ≥ confidence criterion ?
13
Y
N
State answer$_a$
14
Count objects in representation
15
Answer$_a$ = last number in count
16
State answer$_a$
17

(b)

FIG. 5.2. (a) Associative strengths for the strategy choice model. Rows correspond to problems of the form m + n, columns refer to possible answers to problems, and table entries show the associative strength between a given problem and answer. (b) Flow chart for the strategy choice model. "Answer$_a$" refers whichever answer is retrieved on the particular retrieval effort; "problem-answer$_a$ associative strength" refers to the association between the elaborated representation and the retrieved answer (from Siegler & Shrager 1984).

training, and responded to problem probes according to the current distributions.

The model first attempts a direct retrieval. If the associative strength is not sufficiently high, it goes on to the later stages in which some additional representation (such as putting up fingers, or imaging the sets) is used. The model is very successful at accounting for some otherwise puzzling relations among several measures, including proportion of overt strategy use, error rates, and mean solution times. Figure 5.2 depicts the model as a flow-chart. Its computational implementation was written in a programming language that was simply a straightforward conversion of the flow chart. That is, the computational model was not constrained by any issues of working memory, a recognize-act cycle, or the structure of semantic memory.

Implicit Architectures

Soft-core information-processing approaches are not very explicit about the structures or processes that are involved in thinking and development. Case's (1985, 1986) theory of cognitive development illustrates the use of what I call *implicit architectures*. Case postulated *figurative schemes, state representations, problem representations, goals, executive control structures,* and *strategies* in order to account for performance at specific levels of development, and *search, evaluation, retagging,* and *consolidation* to account for development from one performance level to the next. More recently, he has suggested that children can overcome the limitations of short-term memory by acquiring "central conceptual structures" (Case & Griffin, 1990).

Although Case made no explicit reference to symbol structures, his central theoretical construct—what he called Short-Term Storage Space (STSS)—implies that what occupies this space are symbols and symbol structures. Furthermore, the STSS construct assumes that a limited-capacity bottleneck in both storage capacity and computational power accounts for the characteristic differences that many theorists associate with distinct stages of cognitive development. In summary, Case's theoretical constructs appear to require the further assumption of a limited capacity computational architecture that funnels its computational results through the STSS.

Typical of soft-core approaches to information processing in children are models that focus on the structure of thought, without explicit attention to the nature of a computational system that could support that abstract structure. Such approaches, best exemplified by Piaget, have been recently refined and extended by such theorists as Halford (1975) and Fischer (1980). For example, Fischer's skill theory is cast entirely in terms of abstract structures with scant attention to processes. The transition processes that he does discuss—substitution, focusing, compounding, differentiation, and intercoordination—are presented in terms of their global characteristics, and are not constrained by any explicit architecture.

COGNITIVE DEVELOPMENT OCCURS VIA SELF-MODIFICATION OF THE INFORMATION-PROCESSING SYSTEM

Regardless of its location on the hard-soft dimension, every information-processing approach is predicated on the assumption that cognitive development can be characterized as self-modification. This includes accounts ranging from Piaget's original assertions about assimilation, accommodation, and the active construction of the environment, to proposals for various kinds of structural reorganizations (e.g., Case, 1986; Fischer, 1980; Halford, 1970; Kuhn, this volume), to interaction between performance and learning (Siegler, 1987), to explicit mechanisms for self-modifying computer models (Klahr, Langely, & Neches, 1987; Simon, Newell, & Klahr, 1991). This emphasis on self-modification does not deny the importance of external influences such as direct instruction, modeling, and the social context of learning and development. However, it underscores the fact that whatever the external context, the information-processing system itself must ultimately encode, store, index, and process that context. Here too, soft-core approaches tend to leave this somewhat vague and implicit, whereas hard-core approaches make specific proposals about some or all of these processes. However, all information-processing approaches to development acknowledge the fundamental importance of the capacity for self-modification.

Hard-Core Approaches to Self-Modification

In discussing self-modification, I do not make a distinction between learning and development. Instead, I use the more neutral term *change*. (See Klahr, 1989, for a discussion of whether self-modifying production systems are best thought of as models of learning or of development.) It will be understood that change is imposed by the system's own information-processing mechanisms (hence *self-modification*). *Learning* is usually defined as "the improvement of performance over time," but such monotonicity is not assumed here. Indeed, in many areas of development, the measured trajectory is U-shaped, rather than monotone (Strauss, 1982), and a theory of change must ultimately account for these cases.

Many general principles for change have been proposed in the developmental literature. These include: equilibration, encoding, efficiency, redundancy elimination, search reduction, self-regulation, consistency detection, and so on. However, they are not computational mechanisms. That is, they do not include a specification of how information is encoded, stored, accessed, and modified. It is one thing to assert that the cognitive system seeks to avoid unnecessary processing; it is quite another to formulate a computational model that actually does so.

Adoption of a production system architecture allows one to pose focused questions about how broad principles might be implemented as specific mechanisms. One way to do this is to assume the role of a designer of a self-modifying production system, and consider the issues that must be resolved in order to produce a theory of self-modification based on the production-system architecture. The two primary questions are:

1. What are the basic change mechanisms that lead to new productions? Examples are generalization, discrimination, composition, proceduralization, and strengthening.
2. What are the conditions under which these change mechanisms are evoked: when an error is noted, when a rule is applied, when a goal is achieved, or when a pattern is detected?

The recognize–act cycle offers three points at which change can have an effect: A production system's repertoire of behaviors can be changed by affecting the outcome of (a) production matching, (b) conflict resolution, and (c) production application. Each of these is discussed in detail in Neches, Langley, and Klahr (1987), and they are summarized here.

1. Change during the match. The most commonly used technique for altering the set of applicable productions found by the matching process is to add new productions to the set. One way to generate the new productions is to modify the conditions of existing rules. Anderson, Kline, and Beasley (1978) were the first to modify production system models of human learning via *generalization* and *discrimination*. The first mechanism creates a new rule (or modifies an existing one) so that it is more general than an existing rule, while retaining the same actions. The second mechanism—discrimination—creates a new rule (or modifies an existing one) so that it is less general than an existing rule, while still retaining the same actions. The two mechanisms lead to opposite results, although in most models they are not inverses in terms of the conditions under which they are evoked.

2. Change during conflict resolution. Once a set of matching rule instantiations has been found, a production-system architecture still must make some determination about which instantiation(s) in that set will be executed. Thus, conflict resolution offers another decision point in the

recognize–act cycle where the behavior of the system can be affected.

The knowledge represented in a new production is essentially a hypothesis about the correctness of that production. A self-modifying system must maintain a balance between the need for feedback obtained by trying new productions and the need for stable performance obtained by relying on those productions that have proven themselves successful. This means that the system must distinguish between rule *applicability* and rule *desirability*, and be able to alter its selections as it discovers more about desirability. Production systems have embodied a number of schemes for performing conflict resolution, ranging from simple fixed orderings on the rules, to various forms of weights or strengths, to complex schemes that are not uniform across the entire set of productions, to no resolution at all.

3. Changing conditions and actions. Various change mechanisms have been proposed that lead to rules with new conditions and actions. *Composition* was originally proposed by Lewis (1978) to account for speedup as the result of practice. This method combines two or more rules into a new rule with the conditions and actions of the component rules. However, conditions that are guaranteed to be met by one of the actions are not included. For instance, composition of the two rules: AB → CD and DE → F would produce the rule ABE → CDF.

Another mechanism for creating new rules is *proceduralization* (Neves & Anderson, 1981). This involves constructing a very specific version of some general rule, based on some instantiation of the rule that has been applied. This method can be viewed as a form of discrimination learning because it generates more specific variants of an existing rule. However, the conditions for application tend to be quite different, and the use to which these methods have been put have quite different flavors. For instance, discrimination has been used almost exclusively to account for reducing search or eliminating errors, whereas proceduralization has been used to account for speedup effects and automatization.

A basic mechanism for change via chunking was initially proposed by Rosenbloom and Newell (1982, 1987) and first used to explain the power law of practice (the time to perform a task decreases as a power-law function of the number of times the task has been performed). The learning curves produced by their model are quite similar to those observed in a broad range of learning tasks. The chunking mechanism and the production-system architecture to support it has evolved into a major theoretical statement about the nature of the human cognitive system. The system (called Soar) represents the most fully elaborated candidate for a complete cognitive theory—a "unified theory of cognition" as Newell (1990) calls it. It would require a substantial extension of the present chapter to give a comprehensive overview of Soar. However, because the Soar architecture has been used in a recently developed theory of conservation acquisition to be described later, I briefly summarize its main features here.

The Soar architecture is based on formulating all goal-oriented behavior as search in problem spaces. A problem space consists of a set of states and a set of operators that move between states. A goal is formulated as the task of reaching one of a desired set of states from a specified initial state. Under conditions of perfect knowledge, satisfying a goal involves starting at the initial state and applying a sequence of operators that result in a desired state being generated. Knowledge is represented as productions. When knowledge is not perfect, the system may not know how to proceed. For example, it may not know which of a set of operators should be applied to the current state. When such an impasse occurs, Soar automatically generates a subgoal to resolve the impasse. These subgoals are themselves processed in additional problem spaces, possibly leading to further impasses. The overall structure is one of a hierarchy of goals, with an associated hierarchy of problem spaces. When a goal is terminated, the problem solving that occurred within the goal is summarized in new productions called chunks. If a situation similar to the one that created the chunk ever occurs again, the chunk fires to prevent any impasse, leading to more efficient problem solving.

Soar contains one assumption that is both parsimonious and radical. It is that all change is produced by a single mechanism: chunking. The chunking mechanism forms productions out of the elements that led to the most recent goal achievement. What was at first a search through a hierarchy of subgoals becomes, after chunking, a single production that eliminates any future search under the same conditions. Chunking is built into the Soar architecture as an integral part of the production cycle. It is in continual operation during performance – there is no place at which the performance productions are suspended so that a set of chunking productions can fire. Chunking occurs at all levels of sub-goaling, and in all problem spaces. (Soar operates entirely through search in problem spaces: Spaces for encoding the environment, for applying operators, for selecting operators, etc.) Chunking reduces processing by extending the knowledge base of the system.

Simon et al. (1991) used Soar as the theoretical context in which to formulate a computation model of how children acquire number conservation. Their model, called Q-Soar, simulates a training study (Gelman, 1982) in which 3- and 4-year-old children were given a brief training session that was sufficient to move them from the classical nonconserving behavior to the ability to conserve small and large numbers. Q-Soar is designed to satisfy several desirable features of computational models of cognitive development: (a) It is based on a principled cognitive architecture (in this case Newell's Soar theory of cognition); (b) It is constrained by general regularities in the large empirical literature on number conservation; (c) It generates the same behavior as do the children in the specific training study being modeled. That is, it starts out by being unable to pass number conservation tasks, and then, based on the chunks that it forms during the training study, it is able to pass post-tests that include both small and large number conservation tests.

> Q-SOAR's design presumes that young children acquire number conservation knowledge by measurement and comparison of values to determine the effects of transformations on small collections of discrete objects. Having been shown a transformation to a set of objects, the child first categorizes the transformation and then initiates a conservation judgment about the transformation's effect. Ideally, categorization will identify the observed transformation as an instance of a larger class, with effects that are known to be associated (through chunking) with this class. If not, then pre- and post-transformation values created by measurement processes are compared to determine the effect of the transformation. The learning over this processing creates new knowledge about this kind of transformation, which will become available on future occurrences in similar contexts. Now the transformation's effects can be stated without the need for any empirical processing. In other words, the necessity of the effects is recognized. (Simon et al., 1991, p. 438)

Are Other Mechanisms Necessary? Although these processes—generalizations, discrimination, composition, proceduralization, and chunking—may be necessary components of a computational change theory, they may not be sufficient. It is not yet clear whether they could account for the observed differences between the strategies employed by experts and novices (Hunter, 1968; Larkin, 1981; Lewis, 1981; Simon & Simon, 1978). The reorganization necessary to get from novice to expert level may involve much more than refinements in the rules governing when suboperations are performed. Such refinements could presumably be produced by generalization and discrimination mechanisms. However, producing a new procedure requires the introduction of new operations. Those new operations may require the introduction of novel elements or goals—something that generalization, discrimination, and composition and chunking are not clearly able to do.

There are only two simulation studies in which change sequences, and the intermediate procedures produced within them, have been directly observed. Fortunately, a similar picture emerges from both studies. Anzai and Simon (1979) examined a subject solving and re-solving a five-disk Tower of Hanoi puzzle.[5] They found a number of changes in procedure that seemed to require more than the processes listed previously. These included eliminating moves that produced returns to previously visited problem states, establishing subgoals to perform actions that eliminated barriers to desired actions, and transforming partially specified goals (e.g., moving a disk off a peg) into fully specified goals (e.g., moving the disk from the peg to a specific other peg).

---

[5] This puzzle—widely used in psychological studies of problem solving—consists of a "pyramid" of N disks stacked on one of three pegs. The disks are graduated in size, with the

Neches (1981) traced procedure development in the command sequences issued by expert users of a computer graphics editing system. He found a number of changes that involved reordering operations and replanning procedure segments on the basis of efficiency considerations. Subjects were able to evaluate their own efficiency at accomplishing goals and to invent new procedures to reach the same goals more efficiently. Based on these studies, as well as his observations of children "inventing" a novel and efficient strategy for doing simple addition, Neches (1987) created a self-modifying production system called HPM (for Heuristic Procedure Modification). Although these studies just cited deal with adult subjects, the self-modifying processes used by the adults are very likely to be some of the same ones involved in developmental changes.

Metacognition and Information-Processing Approaches

The important point in these examples is that change appears to involve reasoning on the basis of knowledge about the structure of procedures in general, and the semantics of a given procedure in particular. In each example, procedures were modified through the construction of novel elements rather than through simple deletions, additions, or combinations of existing elements.

Beyond the hard-core approaches, this kind of self-analysis of procedures and their byproducts is usually treated as an issue of "metacognition." As Kuhn (this volume) points out, hard-core information-processing approaches have not had a lot to say about metacognition, or reflection, or consciousness, at least not directly. However, HPM is one clear instantiation of the notion, and it is also captured to some extent the way that Soar forms chunks out of the goal trace and local context for satisfied subgoals.

Perhaps the most elaborate consideration of metacognitive processing appears in the "time line processing" sketched by Wallace, Klahr, and Bluff

---

largest disk on the bottom of the stack. The goal is to move the stack from the initial peg to a goal peg, subject to two constraints: (a) only one disk can be moved at a time; (b) a larger disk can never be placed on a smaller disk. The minimum number of moves required to move the N-disk stack from one peg to another is $2^N - 1$. Thus a 5-disk problem requires a minimum of 31 moves. A three-disk version, adapted for use with preschoolers by Klahr and Robinson (1981) is described later in this chapter.

(1987). Wallace et al. described a plan for a self-modifying production system called BAIRN (a Scottish term for child) in which a continuous record of the initial and final conditions of production firings is kept in a time line that provides a sequential record of processing activity. BAIRN learns about the world by processing the information in the time line.

> At first, this information is very primitive, being based [on] only the results of BAIRN's innate endowment of primitive perceptual and motor nodes. As . . . [BAIRN's knowledge of the world] gets more elaborated, so does the information available for processing the time line, as richer and more powerful nodes are added to long-term memory. (Wallace et al., 1987, pp. 360–361)

The general idea in systems like BAIRN, Soar, or HPM is to include mechanisms that enable the system to improve its performance by accessing information about the context and effectiveness of its earlier performance. These mechanisms allow the systems to exhibit the type of behavior that, when seen in humans, is classified as metacognitive.

To the best of my knowledge, only one developmentalist has made the explicit mapping between an information-processing model and metacognition. Siegler's (1989) model of children's strategy choice in multiplication exhibits the emergent property of a rational choice of an efficient and effective strategy. In that model, strategy choices about whether to retrieve the answer to a problem from memory or to calculate the result are made without any rational calculation of the advantages and disadvantages of each strategy. One of the most important features of Siegler's model is that:

> it indicates in detail how a self-regulatory process could operate. The need for self-regulatory processes—executive processes, metacomponents, autonomous regulation, and so on has been persuasively argued previously, but the way in which they accomplish their function has not been clearly elaborated. The present mechanism both resembles and differs from previous suggestions. The mechanism resembles Piagetian and Vygotskian suggestions in that the child's own activity determines future strategy choices. It differs from these and numerous other approaches, however, in that the self-regulation does not depend on reflection or on any other separate governmental process. Instead, it is part and parcel of the system's basic retrieval mechanism. (Siegler, 1988, p. 272)

Summary: Production Systems as Frameworks for Cognitive Developmental Theory

In this section I provide a brief overview of issues that arise in applying production-system architectures to the areas of learning and development.

The framework rests on three fundamental premises of the hard-core approach:

1. The structure of production-system architectures provides insight into the nature of the human information-processing system architecture. This premise derives from observations about similarities in terms of both structural organization and behavioral properties. Structurally, production systems provide a plausible characterization of the relations between long-term memory and working memory, and about the interaction between procedural and declarative knowledge. Behaviorally, strong analogies can be seen between humans and production systems with respect to their abilities to mix goal-driven and event-driven processes, and with their tendency to process information in parallel at the recognition level and serially at higher cognitive levels.
2. Change is a fundamental aspect of intelligence; we cannot say that we fully understand cognition until we have a model that accounts for its development. The first 20 years of information-processing psychology devoted scant attention to the problems of how to represent change processes, other than to place them on an agenda for future work. Indeed, almost all of the information-processing approaches to developmental issues followed the two-step strategy outlined in the Simon quotation that opened this chapter: First, construct the performance model, and then follow it with a change model that operates on the performance model. In recent years, as researchers have finally started to work seriously on the change process, they have begun to formulate models that inextricably link performance and change. Self-modifying production systems are one such example of this linkage.
3. All information-processing system architectures, whether human or artificial, must obey certain constraints in order to facilitate change. It is these constraints that give rise to the seemingly complex particulars of individual production-system architectures. Thus, following from our second premise, an understanding of production-system models of change is a step toward understanding the nature of human development and learning.

The Computer's Role in Simulation Models

Given the centrality of computer simulation to hard-core information processing, it may be useful to address a few common misunderstandings about the role of the computer in psychological theory. First of all, it is important to distinguish between the theoretical content of a program that runs on a computer and the psychological relevance of the computer itself. Hard-core information-processing theories are sufficiently complex that it is necessary to run them on a computer in order to explore their implications. However, this does not imply that the theory bears any resemblance to the computer on which it runs. Meteorologists who run computer simulations of hurricanes do not believe that the atmosphere works like a computer. Furthermore, the same theory could be implemented on computers having radically different underlying architectures and mechanisms. Failure to make the distinction between theory and computer leads to the common misconception that information-processing approaches can be arranged along a dimension of "how seriously they take the computer as a model" (Miller, 1983, p. 250). It would be counterproductive to take the computer at all seriously as a model for cognitive development, because the underlying computer does not undergo the necessary self-modification.

The first computer simulations of developmental phenomena were intended to explain distinct performance levels along a developmental trajectory. These early models did not contain any self-modification mechanisms. Instead, they were intended to explicate the complex requirements for a self-modifying system (an explication entirely absent from Genevan accounts of equilibration). Critics of these early simulation models (Beilin, 1983; Brown, 1982) faulted them for their lack of attention to issues of transition and change. However, the critics failed to appreciate the principal virtue of computational models of distinct developmental levels: that they sharpened the question of self-modification in a way that is simply unattainable in more traditional verbal formulations of developmental theories. In the past few years, several self-modifying systems have been created. These systems—some of which were described in the previous section—exhibit the same performance that, when observed in humans, has been labeled as either learning or development.

A similar misunderstanding of the role of the computer in hard-core information-processing models may have lead to Brown's (1982) widely quoted (but misdirected) criticism that "A system that cannot grow, or show adaptive modification to a changing environment, is a strange metaphor for human thought processes which are constantly changing over the life span of an individual" (p. 100). I agree, but as evidenced by the systems described earlier, the criticism does not apply here: we have some hard-core information-processing approaches that propose very explicit mechanisms for "adaptive modification to a changing environment."

The hard-core information-processing approaches are serious, not about the similarity between humans and computers, but rather about the extent to which intelligent behavior—and its development—can be accounted for by a symbol-processing device that is manifested in the physical world. The strong postulate for hard-core information-processing is that both computers and humans are members of the class of "physical symbol systems" (Newell, 1980), and that some of the theoretical constructs and insights that have come out of computer science are relevant for cognitive developmental theory. One such insight is what Palmer and Kimchi (1986) call the *recursive decomposition* assumption: Any nonprimitive process can be specified more fully at a lower level by decomposing it into a set of subcomponents and specifying the temporal and informational flows among the subcomponents. This is a good example of how abstract ideas from computer science have contributed to hard-core information processing: "It is one of the foundation stones of computer science that a relatively small set of elementary processes suffices to produce the full generality of information processing" (Newell & Simon, 1972, p. 29). An important consequence of decomposition is that

> . . . the resulting component operations are not only quantitatively simpler than the initial one, but *qualitatively different* from it. . . . Thus we see that higher level information-processing descriptions sometimes contain *emergent properties* that lower level descriptions do not. It is the *organization* of the system specified by the flow relations among the lower level components that gives rise to these properties. (Palmer & Kimchi, 1986, pp. 52–53)

The importance of emergent properties cannot be overemphasized, for it provides the only route to explaining how intelligence—be it in humans or machines—can be exhibited by systems comprised of unintelligent underlying components—be they synapses or silicon. Even if one defines underlying components at a much higher level—such as production systems or networks of activated nodes, emergent properties still emerge, for that is the nature of complex systems.

The emergent property notion provides the key to my belief that hard-core information-processing approaches provide a general framework, particular concepts, and formal languages that make possible the formulation of powerful theories of cognitive development. The fundamental challenge is to account for the emergence of intelligence. Intelligence must develop from the innate kernel. The intelligence in the kernel, and in its self-modification processes, will be an emergent property of the *organization* of elementary (unintelligent) mechanisms for performance, learning, and development. Thus, the issue is not "sacrificing explanatory breadth for explanatory precision" (Kuhn, this volume), but rather achieving explanatory breadth on the basis of the emergent properties revealed by explanatory precision.

USING HIGHLY DETAILED ANALYSES OF THE ENVIRONMENT FACING THE CHILD ON SPECIFIC TASKS

The realization that investigation of psychological processes presupposes a highly developed, abstract analysis of the task and available constraints has

> perhaps been the major advance in psychology in the last several decades. (Kellman, 1988, p. 268)

Kellman's observation echoes Simon's (1969, Chap. 2) well-known parable of the ant, whose complex path toward a goal was characterized as a set of simple mechanisms encountering a complex and irregular environment. Simon's claim was that, in the human as well as in the ant, much of the apparent complexity of behavior is a function of the complexity of the environment rather than of the cognitive system. This insight is particularly important for developmentalists, for it demands that our explanations for changes in behavior include an account of changes in both the organism and the environment in which it is embedded.

All of the methodological practices to be described in the remainder of this chapter start with a careful task analysis. Both chronometric techniques and error analysis require at least a rudimentary analysis of the task environment. In addition, there are some information-processing approaches in which complex and detailed task analysis plays a central role, even when neither error analysis nor chronometrics are used. In a sense, these approaches consist of nothing but task analysis. While such work is typically preliminary to further work in either error analysis or computer simulation (or both), it is often useful for its own sake, as it clarifies the nature of the tasks facing children.

Klahr and Wallace's (1970b) task analysis of class inclusion is an early example of such a formal characterization of an important developmental task. Their goal was to illustrate how a common "Piagetian experimental task" (i.e., the full set of test items that are typically given when assessing class inclusion competence, including finding some objects, finding all objects, comparing subsets of objects, etc.) involved the coordination of several more basic information processes. They proposed a network of interrelated processes—similar to Gagne's (1968) learning hierarchies—in which some processes had common subcomponents, while others were relatively independent. Klahr and Wallace's analysis enabled them to explain how surface variations in a task could invoke different processes, that, in turn, would have profound effects on performance, even though the underlying formal logic of the task remained invariant.

In the area of children's counting, Greeno, Riley, and Gelman (1984) formulated a model for characterizing children's competence. Their model is much more complex than the early Klahr and Wallace analysis of classification, but it is fundamentally similar with respect to being a formal task analysis whose primary goal is to elucidate the relations among a set of underlying components.

Klahr and Carver (1988) used a formal task analysis to design an instructional unit to teach elementary school children how to debug computer programs. The unit was designed to be inserted in the normal curriculum for teaching a graphics programming language. Children tend to write programs that are "buggy," that is, programs in which the desired picture does not match the picture drawn by the child's program. Children typically fail to acquire very effective procedures for debugging programs, so Klahr and Carver attempted to teach the necessary skills explicitly. First they analyzed the components of debugging into four distinct phases.

1. Bug identification—the child generates a description of the discrepancy between the program plan (e.g., what the desired picture should look like) and the program output (e.g., what the program actually drew). Based on the discrepancy description, propose specific types of bugs that might be responsible for the discrepancy.
2. Program representation—the child articulates the structure of the program in order to investigate the probable location of the buggy command in the program listing.
3. Bug location—the child uses the cues gathered in the first two phases to examine the program in order to locate the alleged bug.
4. Bug correction—the child examines the program plan to determine the appropriate correction, replaces the bug with the correction in the program, and then reevaluates the program.

Based on the formal task analysis, Klahr and Carver then created a production system model that could actually do the debugging, and they used the productions in the model to specify a set of cognitive objectives for insertion in a programming curriculum (Carver, 1986). In addition to the instructional elements, their debugging model provided a framework for assessment of debugging skills, for creation of transfer tasks, and for evaluation of transfer. Thus, the entire instructional intervention (which was very successful in teaching debugging skills) was based on the initial task analysis.

USING FORMAL NOTATIONAL SCHEMES FOR EXPRESSING COMPLEX, DYNAMIC SYSTEMS

The use of computer-simulation languages is the sine qua non of hard-core information processing. Nevertheless, there are several lesser degrees of formalization that mark the soft-core methods, including such devices as scripts, frames, flow charts, tree diagrams, and pseudo-programming languages. The attractive property of any of these formal notations is that they tend to render explicit what may have only been implicit, and they frequently eliminate buried inconsistencies. That is, compared to verbal statements of theoretical concepts and mechanisms, each of these notations offers increased precision and decreased ambiguity.

Flow charts are perhaps the most common type of formal notation used by information-processing psychologists. For example, Sternberg and Rifkin (1979) used a single flow chart to represent four distinct models of analogical reasoning. Their depiction clearly indicates how the models are related and what parameters are associated with each component of each model.

Another type of formal notation commonly used in research on children's comprehension of stories is the *story grammar* (Mandler & Johnson, 1977; Stein & Glenn, 1979). Nelson has analyzed children's event representations in terms of *scripts* (Nelson & Gruendel, 1981). Mandler (1983) provides a comprehensive summary of how these kinds of representations have been used in developmental theory. In both areas, the underlying theoretical construct is the *schema*: an organized knowledge structure containing both fixed and variable components. The fixed components bear relations that are characteristic of the general properties of the situation represented by the schema, and the variable components represent the specific instance that is currently being processed. For example, a story grammar would have components for the main protagonist, the goal or intent of the protagonist, an obstacle or threat to the achievement of the goal, and the resolution of the threat. For event representations, children appear to have scripts for common activities such as going to a restaurant, in which the fixed components include driving, parking, ordering, eating, and paying, and the variable components might include the order in which the events occur (e.g., pay before or after eating), the particular things ordered, the seating arrangement, and so on.

As with any other of the constructs used in information-processing approaches, the schema construct can be used in a variety of ways, and with varying degrees of ambiguity (Mackworth, 1987). However, it is possible to be quite specific about what one means by the term. For example, Hill and Arbib (1984) attempted to clarify some of the different senses in which the term *schema* has been used, and they go on to describe a schema-based computational model of language acquisition.

Given this range of notational options for describing information-processing theories, what criteria should be used in choosing among them? This issue is discussed at length by Klahr and Siegler (1978). They suggest that the following four criteria be used in choosing a representation:

1. Is the representation sufficient to account for behavior? Does it have a clear mapping onto the empirical base for which it is supposed to account?

2. Is the representation amenable to multiple-level analyses? Is it easy to aggregate and disaggregate the grain of explanation? That is, can one easily go from a characterization of the average behavior of a group of children to more specific models that capture individual performance? For the design of well-controlled experiments or curriculum design, the representation will have to be stated in terms of averages across many subjects; it must be a modal form. For detailed study of individual strategies and component processes, it must be capable of disaggregation without drastic revision.
3. Is the representation consistent with well-established processing constraints?
4. Does the representation have "developmental tractability" (Klahr & Wallace, 1970a)? That is, does it allow the theorist to state both early and later forms of competence and provide an easy interpretation of each model as both a precursor and successor of other models in a developmental sequence?

What about mathematical models of developmental phenomena? Should they be included in the set of formal notational schemes that signal soft-core information processing? The situation is not straightforward. On the one hand, mathematical modeling certainly meets the criteria of formalization and precision. Indeed, the following argument for mathematical models could equally well be made for computational models.

> It is precisely because the phenomena are so complex that we must have mathematics. Even in relatively simple (one might suppose) areas of psychology, a reader of the literature can easily be led down the primrose path through verbal argument. The logic seems impeccable. However, when the psychological principles on which the theory is based are put into mathematical form, the stated predictions may fail to follow at all. Moreover, just as in other sciences, the predictions are often rendered more testable by being derived as mathematical propositions or theorems. (Townsend & Kadlec, 1990, p. 227)

Nevertheless, most of the developmentally relevant mathematical modeling has focused on perception, rather than cognition. Those models that have addressed higher order cognitive developmental issues have characterized information processing at a very abstract level: in terms of states and transition probabilities, rather than in terms of structural organization and processes that operate on that structure (e.g., Brainerd's, 1987, Markov models of memory processes). As Gregg and Simon (1967) demonstrated very clearly with respect to stochastic models of concept learning, most of the interesting psychological assumptions in such models are buried in the text surrounding the mathematics. They point out that "the accurate predictions of fine-grain statistics that have been achieved with [stochastic theories] must be interpreted as validations of the laws of probability rather than of the psychological assumptions of the theories" (p. 275).

To cite a specific example of this general problem from the developmental literature, Wilkinson and Haines (1987) used Markov learning models to propose some novel answers to the important question of how children assemble simple component skills into reliable strategies. However, they couched their analysis in terms of the probabilities of moving between abstract states, while their discussion in the text was rife with undefined processes whereby the child "discovers," "adopts," "retains," "invokes," "moves," "prefers," "abandons," or "reverts." As is often the case in the use of mathematical models, the formalism of the mathematics obscures the informality of the underlying theory. Perhaps this is the reason why mathematical modeling has not played a central role in information-processing approaches to cognitive development.

MEASURING THE TIME-COURSE OF COGNITIVE PROCESSING

Many information-processing psychology studies of children's thinking ask questions about the rates at which different mental processes occur. When the mental processes of interest have durations of seconds or fractions of seconds, the methodology associated with their analysis is called *chronometric analysis*. The focus in these studies is on how a specific mental algorithm or strategy is organized and executed. When the focus shifts from how these strategies work to where they came from in the first place, it becomes necessary to study children's performance repeatedly over several days or weeks or perhaps months, seeking characteristic patterns that signal changes in the organization and content of underlying processes. Medium-duration studies of this type are called *microgenetic studies*. In the next two sections I describe each kind of methodology.

Chronometric Analysis

Chronometric analysis is based on three assumptions. First, there is a set of distinct, separable processes that underlie the behavior under investigation. Second, the particular process of interest can be isolated, via a task analysis, such that experimental manipulations can systematically induce the system to increase or decrease the number of executions of the focal process. The third assumption is that the experimental manipulations affect *only* the number of executions of the focal process, and nothing else about that process or the total set of processes in which it is embedded. (For a thorough discussion of the history and methodology of chronometric studies, primarily with adults, see Chase, 1978.)

Chronometric analysis can be used at several levels of aggregation. At the smallest grain sizes, it is used to obtain estimates of the mean time to execute underlying processes. At larger grain sizes, it is used to determine the overall organization of a cognitive process comprised of smaller components. In the following descriptions, I start with examples of the finer grained use of chronometric analysis. Then I describe a few examples of chronometrics at a more aggregate level.

Memory Scanning. The use of chronometric methods with children is exemplified by Keating and Bobbit's (1978) extension of Sternberg's (1966) memory-scanning paradigm. The question of interest here is how people search their short-term memory. The basic procedure is to present subjects with a set of several digits, followed by a *probe* digit. The subject's task is to decide whether the probe digit was in the original set. The main independent variable is the size of the original set. Reaction time is measured from the onset of the probe until the subject responds. In addition to the general assumptions listed above, the paradigm assumes that the items in the set are stored in some kind of passive buffer, and that there is an active process that sequentially compares the probe with each of the items stored in the buffer. The empirical question is how long each comparison (and move to the next item) takes. Sternberg had discovered that when adults were attempting to decide whether a probe item was a member of a previously stored list, they could compare the probe item to the list items at the rate of approximately 20 items per sec (or 50 msec per item). Furthermore, adults appear to use an exhaustive search: They go through the entire list regardless of whether or not a match is found along the way. Keating and Bobbit found that 9-year-olds took almost twice as long per item as 17-year-olds.

Other Basic Cognitive Processes. Such age differences in processing rates are found in almost all chronometric studies. Indeed, as Kail (1991b) noted in his review of 72 studies comparing processing speed in children and adults: "Age differences in performance on speeded tasks are large and remarkably consistent" (p. 490). Kail (1988) suggested the following explanation for these differences:

> One hypothesis is that age differences in processing time reflects changes that are specific to particular processes, tasks, or domains. For example, age differences in processing speed may reflect the developmental acquisition of more efficient strategies for task solution. . . . A second hypothesis is that age

> differences in processing speed are due to more general developmental change. For example, in information-processing theories, performance on many cognitive tasks requires processing resources or attention. . . . increasing resources typically increases speed of processing, even when all other factors are held constant. Therefore, age-related increases in the amount of processing resources could produce age-related increases in processing speed. (pp. 339–340)

Kail (1988) reasoned that one could distinguish between the two hypotheses by examining the general shape of the functions that plot age versus processing speed for a variety of tasks. "Specifically, if some central mechanism changes monotonically with age, and if the function that relates decreases in processing time to changes in this central mechanism has the same form for two or more processes, the form of the growth function should be the same for those processes" (p. 340). His work represents an elegant example of the extent to which chronometric analysis can illuminate important developmental questions.

For each of the 15 ages from 8 years to 22 years (e.g., 8-year-olds, 9-year-olds, and so forth), Kail estimated the processing rate for five different tasks that involve very basic mental processes. For each task, Kail arranged the stimulus materials so that the process in question had to be executed repeatedly as a function of the stimulus. This enabled him to estimate the duration of the underlying focal process. The five tasks, and the resulting rates were:

1. Visual search. The stimuli were the digits 1 to 9. First a single digit—the *study digit*—appeared on a computer screen. Then, after a short delay, a set of one to five digits appeared. This was the *probe set*. The subject's task was to signal, as fast as possible, whether or not the probe set contained the study digit. Note that in this task the subject had to match a digit from memory (the study digit) with each of the digits in an external display. The processing time per item ranged from about 80 msec for the 8-year-olds to about 25 msec for the adults. (See Fig. 5.3 for the results from this and the other four Kail tasks.)
2. Memory search. Here Kail used the standard Sternberg (1966) memory scanning paradigm described earlier, with the same kind of materials as in the Visual Search task. For this task, each trial started with the subject learning a set of digits (set size 1, 3, or 5). Once the initial set had been studied, a probe digit was presented, and the subject had to indicate if the probe was a member of the study set. Here, the single digit in the external display had to be matched against a mental representation of the study set. The processing time per item ranged from about 150 msec for the 8-year-olds to about 50 msec for the adults.

3. Mental rotation. In this task, subjects were presented with a pair of letters in any of six different orientations, and they had to decide whether the letters were identical or mirror images. At all ages, reaction time increased with increasing orientation, and the slope of the reaction time versus orientation function decreased with age. Mean slopes were 4.5 msec/degree for children, 3.3 for adolescents, and 3.0 for adults.

4. Name retrieval (reported originally in Kail, 1986). The purpose of this task was to estimate the time necessary to retrieve the name of something, given a picture of it. Stimuli were pairs of pictures of common objects in two formats (e.g., an open umbrella or a closed umbrella, a peeled or an unpeeled banana). These pictures were combined into three different kinds of pairings: (a) pairs that were identical physically and in name (e.g., a pair of open umbrellas); (b) pairs that were identical in name only (e.g., an open and a closed umbrella); and (c) pairs that were different both physically and in name (e.g., an umbrella and a banana). Subjects were presented with a series of these different pairs, and given two types of instruction. In one condition, the subjects had to decide whether the pairs of objects had the same name, and in the other they had to decide if they were physically identical. By subtracting the response times on those trials that required the subjects to retrieve the name of the object from the response times on those trials that required only a physical match, Kail was able to estimate the mean name retrieval time for each of the age groups studied. It ranged from approximately 300 msec for the youngest children to about 150 msec for adults.

5. Mental addition. Subjects were presented with problems of the form $m + n = k$, where $1 \leq m, n \leq 9$. Problems with $n = m$ were not used. For half the problems the sum was correct and for the other half it was incorrect. Subjects responded by pressing either of two response buttons. Kail based his analysis of subject's response times on Ashcraft's (1987) associative retrieval model in which solution of these problems involves entering an arithmetic network at nodes corresponding to m and n, then searching for the intersection at which is stored the sum. The model assumes that memory search time increases as a function of the square of the sum, and that overall response time is a function of whether the equation $m + n = k$ is true or false. Accordingly, Kail estimated the memory search rate by using multiple regression to fit the median RTs at each age to the function

$$RT = B(m + n)^2 + t + k$$

where B is the memory search rate, and t is the additional amount of time to respond "false." Memory search rates ranged from approximately 7.5 sec per squared increment for the 8- and 9-year-olds to less than 3 sec for the adults.

Having determined the processing time per item for each age group on each of the five tasks, Kail then determined the relations between processing time and age for each task. Not surprisingly, for all of the tasks, there was a decrease in processing time with age. But more important, the best-fitting function for processing time versus age was an exponential decay function that could be fit by a single decay parameter (see Fig. 5.3). Furthermore, these exponential decay curves are found in speeded perceptual-motor tasks (Kail, 1991a), as well as the cognitive tasks described here. Kail interpreted these results by positing an increasing amount of common, nonspecific processing resources that become available to children as they develop: "Common growth functions are found because the increased resources yield a constant increment in speeded performance across tasks" (p. 362).

Kail's work represents an interesting mix of the hard-soft dimensions that I have been using to characterize the field. With respect to experimental methodology, it is about as hard as it can get. The experiments are very clean and the analysis is deeply quantitative. However, with respect to theoretical assumptions, it is at the soft end of the spectrum: It posits no clear mechanism through which the vague construct of "processing resources" might be realized. Nevertheless, these results provide an important empirical constraint for such models.

In addition to the three standard assumptions—listed earlier in this section—underlying chronometric analysis, Kail's approach is based on a fourth: that the organization of the strategy for accomplishing a task remains constant across ages, and only the speed of processing changes. That is, Kail assumed that children's processes for memory scanning were organized in the same way as adults'. Then he proceeded to estimate some of the critical parameters of these processes and to chart their developmental course. However, in many cases, rather than presuming that the organization is known and age-invariant, the researcher's goal is to determine just what that organization is at different ages or skill levels. Chronometric analysis can be applied to this situation at a somewhat coarser grain size. The focus is not so much on individual processing rates as on the overall temporal pattern of responses generated by different cognitive strategies. The next three examples illustrate this kind of coarser grained use of chronometric analysis.

Mental Arithmetic. One of the best-known studies using chronometric analysis with children was Groen and Parkman's (1972) analysis of how first graders solved simple addition problems. Groen and Parkman proposed several alternative models of how these children might add two single digit numbers to produce their sum. One plausible model would be for the child to represent the first argument, then the second argument, and then *count out* the sum. The actual representation could be external (on fingers or

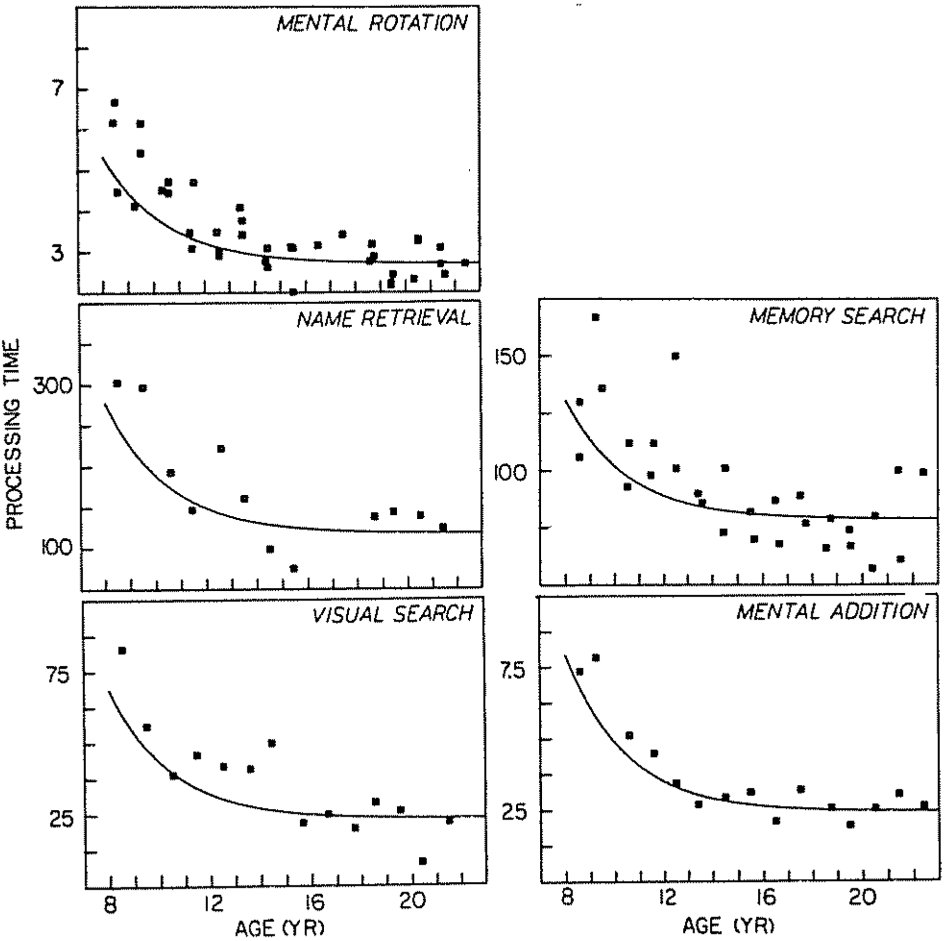


FIG. 5.3. Developmental functions for rates of mental rotation (data from Kail, 1988, Experiment 2 and from Kail, 1986, Experiments 1 and 2), name retrieval (from Kail, 1986, Experiment 1), memory search (from Kail, 1988, Experiments 1 and 2), visual search (from Kail, 1988, Experiment 1), and mental addition (from Kail, 1988, Experiment 2). Rate of mental rotation is estimated by the slope of the function relating response time to the orientation of the stimulus. Name retrieval is estimated by the difference between times for name and physical matching. Visual search is estimated by the slope of the function relating response time to the size of the search set. Memory search is estimated by the slope of the function relating response time to the size of the study set. Retrieval of sums on the mental addition task is estimated from the slope of the function relating response time to the sum squared. The solid line depicts values derived from the best-fitting 11-parameter exponential function (i.e., one in which the decay parameter, *c*, is the same for all five processes; from Kail, 1988).

counting blocks) or internal. In either case, the expected time to compute the sum would be proportional to the sum of the two numbers. Another strategy might be to *count on* from the first number: start with the first, and count as many steps as the second. A child following this strategy on the problem 3 + 5, would, in effect, say to herself "3 + 5; that's 3, 4, 5, 6, 7, 8—the answer is 8." Response times for this strategy would be proportional to the value of the second number in the pair of addends. Yet another strategy is the *min model*. This is like counting on, except that the child always starts counting from the larger of the two arguments, thereby minimizing the number of steps (e.g., for both 3 + 5, and 5 + 3, the child would start with the 5, and increment it 3 times). Times for this strategy are proportional to the minimum of the two arguments.

Reasoning in this way, Groen and Parkman predicted a pattern of reaction times as a function of several relations among the two addends (sum, difference, min, max). Based on their analysis of mean reaction times across subjects and trials, Groen and Parkman concluded that the min model provided the best fit to their data. (Even at the time, there were some exceptions to this general result, and further analysis by Siegler & Jenkins, 1989—described later—revealed a much more complex picture. Nevertheless, the initial Groen and Parkman work still stands as a pioneering effort in chronometric analysis of children's performance.)

Transitive Reasoning. "Bill is taller than Sue and shorter than Sally. Is Sally shorter than Sue?" How do children solve this kind of transitive inference problem? Ever since Burt (1919) first explored children's developing ability to deal with transitive relations, developmentalists have been interested in this question. One specific question that arises in this context is how the individual premises (X is taller than Y) are stored and accessed. There are two distinct possibilities. One alternative is that children construct an integrated representation of the individual items as the pairwise relations are presented. Then, when a probe is presented, they *read off* the relative sizes of the probe items from the integrated display. Another alternative is that children store the individual pairs, and at probe time they link them together to produce the answer.

Trabasso and his colleagues used chronometric analysis to decide the issue (Trabasso, 1975; Trabasso, Riley, & Wilson, 1975). They reasoned that if children use an integrated representation, then their response times should show the same pattern for the internally stored representation as for an external visual display of the same ordered set of objects. In particular, the pattern should exhibit the familiar *symbolic distance* effect, in which it takes less time to determine the ordering for two widely separated objects, than to decide the ordering for two adjacent items. On the other hand, if children are connecting the premises at probe time, then the closer objects should take less time to resolve (because there are fewer connections to make). Trabasso and his colleagues investigated this problem by using six-term problems (the example given at the opening of the previous paragraph was a three-term problem). They presented children with repeated exposure to all adjacent pairs (i.e., AB, BC, CD, DE, EF), in both orders (e.g., A is larger than B; B is smaller than A), appropriately randomized, until they met a learning criterion for each adjacent pair. Then they presented probe questions about all possible pairs and measured response times. The probes included pairs of items (in both orders; i.e., both BF and FB) that were zero inferential steps apart (i.e., the adjacent pairs used in the training lists, such as BC), pairs that were one inferential step apart (e.g., BD), and pairs that were two inferential steps apart (e.g., EB). The subjects included a group of 6-year-olds, a group of 9-year-olds, and a group of adults.

The resulting reaction time patterns were (with minor exceptions) very consistent: Reaction time was inversely related to the number of inferential steps. That is, pairs that were very far apart produced faster responses than the adjacent pairs on which subjects had been trained. This was true at all ages (as expected, older subjects were faster than younger subjects), and for all display conditions (verbally presented pairs, visually presented pairs, and an integrated visual display of all objects.) By using this kind of chronometric analysis in a series of related studies with a variety of subject populations, Trabasso (1975) was able to make some very strong statements about an important mnemonic skill:

> Our analysis suggests that children ranging in age from 4 to 10 years-of-age, mentally retarded adolescents and college students use similar strategies of constructing linear orders from pairwise, ordered information, store this representation in memory and use it to make comparative relations on all members in the array. . . . In short, we believe that we have provided a mechanism for information integration and inference-making that cuts across a variety of situations and tasks. (pp. 167–168)

Elementary Quantification: Subitizing and Counting. Our final example represents a mix of the two kinds of chronometric analysis described here. In this case, the goals of the research are twofold: (a) to determine whether or not children and adults use the same general strategies and (b) to estimate the rates of the components of those strategies. Chi and Klahr (1975) addressed the question of how kindergarten children and adults quantify (i.e., generate an internal quantitative symbol for) displays of discrete objects. One quantification strategy might be simply to count each object. The processing time for counting should be a linear function of the number of objects being counted. On the other hand, the earliest conceptions of the "span of apprehension" (Jevons, 1871) assumed that there was some number, *N*, of discrete objects that the mind could immediately perceive, apprehend, or recognize. Such a process—later called *subitizing* (Jensen, Reese, & Reese, 1950)—would produce a flat slope of reaction time versus *N*. In addition to assessing these two positions, Chi and Klahr addressed the developmental question of how the two processes differed in children and adults.

Subjects were presented with randomly arranged displays of *N* dots and asked to say as rapidly as possible how many there were in the display. The range of *N* was from 1 to 10 for adult subjects and from 1 to 8 for the children. Reaction times were measured from the onset of the display to the beginning of the verbal response. The results are shown in Fig. 5.4. For both the adults and the children, the mean reaction times were best fit by a two-segment linear regression analysis with a break point between $N = 3$ and $N = 4$. For $N \leq 3$—the subitizing range—the slope of the adult function is about 50 msec per dot, whereas for the children it is nearly 4 times as great. For $N \epsilon 4$—the counting range—the slope is about 300 msec for adults and about 1 sec for children. Error rates were nearly zero for $N \leq 4$ for both children and adults. Beyond that range, they abruptly increased to about 25% for children, and about 5% for adults. Based on the characteristic pattern of results (both RTs and errors) and the specific parameter estimates for rates and ranges, Chi and Klahr (1975, p. 438) concluded that in both adults and young children, there appear to be two

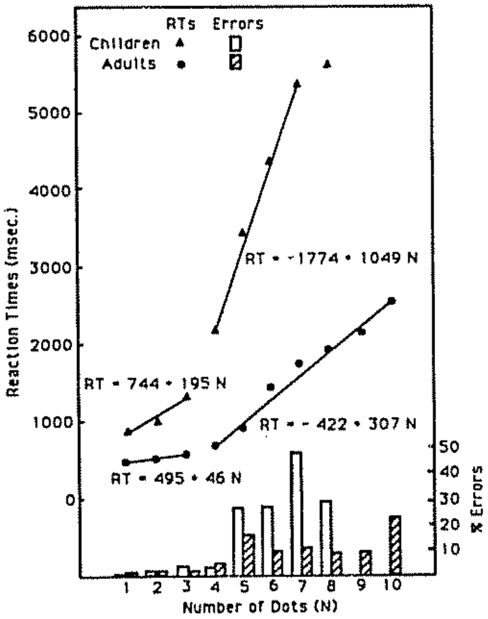


FIG. 5.4. Mean reaction times and error rates for children and adults as a function of the number of dots in the display (from Chi & Klahr, 1975).

distinct quantification processes. One process, operating almost errorlessly on the range below $N = 4$ is 5 to 6 times as rapid as the other, which operates on the range above $N = 3$.

Microgenetic Studies

In the context of chronometric studies, the phrase "time course of cognitive processing" implies brief tasks with components identified at the level of fractions of a second. But the phrase can also refer to the much longer intervals (weeks, months, or years) over which cognitive change occurs. The most common way to investigate change is to design cross-sectional studies in which the same task is presented to groups of subjects at different ages (e.g., Kail's studies described earlier). Somewhat less common are longitudinal studies in which the same group of subjects is assessed repeatedly over an extended time period. Typically, the observation points in longitudinal studies are months or years apart and the measurements are relatively crude when compared to chronometric tasks. However, an interest in the more detailed aspects of changes in children's information processing has led to an approach called the *microgenetic method* that is particularly well suited to detecting changes in children's strategies.

> Three key properties define the microgenetic approach: (a) Observations span the entire period from the beginning of the change of interest to the time at which it reaches a relatively stable state; (b) The density of observations is high relative to the rate of change of the phenomenon; (c) Observed behavior is subjected to intensive trial-by-trial analysis, with the goal of inferring the processes that give rise to both quantitative and qualitative aspects of change. (Siegler & Crowley, 1991, p. 606)

Siegler and Crowley summarize the history and current status of microgenetic studies, and then provide a detailed account of one study that exemplifies the microgenetic method. Siegler and Jenkins (1989) focused on how children discovered the min strategy for addition (described earlier). They followed eight 4- and 5-year-old children over an 11-week period. At the start of the period all of the children were proficient at simple addition (problems with addends 1–5 inclusive), and their most common addition strategy was to count from 1. Children received seven problems in each of approximately three sessions per week during the 11-week period. In order to determine what strategy a child used on each problem, Siegler and Jenkins used a variety of methods such as observing their behavior (counting on fingers, and so forth) and measuring speed and accuracy. However, their primary method involved simply asking children how they solved each problem. The following example is taken from a trial on which a 5-year-old first used the min strategy (Siegler & Crowley, 1991, p. 613):

E: How much is 2 + 5?
S: 2 + 5—(whispers), 6, 7—it's 7.
E: How did you know that?
S: (excitedly) Never counted!
E: You didn't count?
S: Just said it—I just said after 6 something—7, 6—7.

The rich data set produced by this high-density measurement technique yielded a correspondingly rich portrait of developmental change. At the most aggregate level, children improved from about 75% correct to nearly perfect performance over the 11 weeks. More interesting than just the outcome of each trial was the pattern of strategies used to produce those outcomes. Overall, children used half a dozen different strategies, and, more important, this variability was true not only for the group, but for individual children. Furthermore, the study provided clear data on the discovery of new strategies, on the precursors of strategy discovery, and on the subsequent consequences of strategy discovery. Based on their own work and that of others, Siegler and Crowley (1991) conclude that

> . . . microgenetic experiments have yielded closely parallel results across quite diverse changes. One such finding involves the halting and uneven use of newly acquired competencies. Even after children discover sophisticated scientific experimentation strategies, they often continue to use less sophisticated ones as well (Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Phelps, 1982; Schauble, 1990). When they discover a new problem solving method with the help of their mothers, they may later fall back on shared control rather than continuing to exert sole responsibility for its execution (Wertsch & Hickmann, 1987). New concepts about the workings of gears are applied in a similarly sporadic fashion (Metz, 1985), as are new strategies for adding numbers (Siegler & Jenkins, 1989). (p. 618)

Another common finding of microgenetic studies is that innovations occur following successes as well as failures. Discoveries have been found to follow successes, rather than impasses or errors, in many children's map drawing and language use (Karmiloff-Smith, 1984), arithmetic (Siegler & Jenkins, 1989), pictorial representations (Inhelder et al., 1976), and scientific experimentation strategies (Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Phelps, 1982; Schauble, 1990). These findings point to the importance of observing in a variety of domains the frequency and types of variation produced without apparent external motivation.

USING HIGH-DENSITY DATA FROM ERROR PATTERNS AND PROTOCOLS TO INDUCE AND TEST COMPLEX MODELS

Pass/fail data provide only the crudest form of information about underlying processes. Nevertheless, most of the empirical research in cognitive development is reported in terms of percentage of correct answers. Another characteristic of the information-processing approach is the premise that much more can be extracted from an appropriate record of children's performance. The basic assumption is that, given the goal of understanding the processing underlying children's performance, we should use all the means at our disposal to get a glimpse of those processes as they are occurring, and not just when they produce their final output. Verbal protocols, eye movements, and error patterns (as well as chronometric methods, mentioned earlier) all provide this kind of high-density data. Examples of some of these methods have already been provided in previous sections, but here we look at them in more detail.

The view that detailed error analysis provides a powerful window into the child's mental processes is neither novel nor radical. Piaget's pioneering analysis (Piaget, 1928, 1929) of children's characteristic errors and misconceptions in a wide variety of domains made him, in effect, a founding member of the soft-core information-processing club. He was probably the first to demonstrate that children's errors could reveal as much, or more, about their thought processes as their successes, and a substantial proportion of his writing is devoted to informal inferences about the underlying knowledge structures that generate children's misconceptions in many domains (see Kuhn, this volume). Siegler (1981) put the issue this way:

> Many of Piaget's most important insights were derived from examining children's erroneous statements; these frequently revealed the type of changes in reasoning that occur with age. Yet in our efforts to make knowledge-assessment techniques more reliable and more applicable to very young children, we have moved away from this emphasis on erroneous reasoning and also away from detailed analyses of individual children's reasoning. . . . The result may have been a loss of valuable information about the acquisition process. . . . [My] hypothesis is that we might be able to increase considerably our understanding of cognitive growth by devoting more attention to individual children's early, error-prone reasoning. (p. 3)

Analysis of Error Patterns

The basic assumption in error-analytic methodologies is that children's knowledge can be represented as a set of stable procedures that, when probed with an appropriate set of problems, will generate a characteristic profile of responses (including specific types of errors). Application of this idea to children's performance reached perhaps its most elegant form in the computer simulation models of children's subtraction errors by Brown and his colleagues (Brown & Burton, 1978; Brown & VanLehn, 1982). Brown and his colleagues demonstrated that a wide variety of children's subtraction errors could be accounted for by a set of "bugs" in their calculation procedures. For example, two of the most frequent bugs discovered by Brown and Burton were:

BORROW FROM ZERO:

When borrowing from a column whose top digit is 0, the student writes 9, but does not continue borrowing from the column to the left of the zero.

$$\begin{array}{r}103\\-45\\\hline 158\end{array}$$

SMALLER FROM LARGER:

The student subtracts the smaller digit in a column from the larger regardless of which one is on top.

$$\begin{array}{r}254\\-118\\\hline 144\end{array}$$

These and dozens of more subtle and complex bugs were inferred from the analysis of thousands of subtraction test items from 1,300 children. The key to the analysis was the creation of a network of subprocedures that comprise the total knowledge required to solve subtraction problems. This procedural network was then examined for possible points of failure, to explain the patterns of erroneous answers.

Another highly productive research program based on the analysis of error patterns is Siegler's well-known *rule assessment* methodology (Siegler, 1976, 1981). The basic idea in this and other developmentally oriented error-analysis work (e.g., Baylor & Gascon, 1974; Klahr & Robinson, 1981; Young, 1976) is that, at any point in the development of children's knowledge about a domain, their responses are based on what they know at that point, rather than on what they don't know. In order to characterize that (imperfect) knowledge, the theorist attempts to formulate a model of partial knowledge that can generate the full set of responses—both correct and incorrect—in the same pattern as did the child. The model thus becomes a theory of the child's knowledge about the domain at that point in her development.

Fay and Mayer (1987) applied this kind of error analysis to the domain of spatial reference. They were attempting to teach children (from 9 to 13 years old) how to write programs in Logo, a programming language in which children write commands for a "turtle" that draws lines on the computer screen. Logo includes commands that can move the turtle forward or backward (FD or BK) a specified number of units, and that can rotate the turtle to the left or right (LT or RT) a specified number of degrees. For example, to draw a square box 5 units on a side, the child might write the following program:

FD 5 RT 90 FD 5 RT 90 FD 5 RT 90 FD 5 RT 90.

During the execution of this program, the turtle would be in four different orientations, and the RT command would be interpreted relative to the turtle's current orientation, rather than absolutely.

The distinction between relative and absolute orientation is difficult for children of this age, and so Fay and Mayer used the Logo context to study children's naive conceptions about spatial reference. They examined how children interpreted Logo commands to move and turn from various initial orientations. Children were presented with problems that varied in initial orientation of the turtle, the type of command (move or turn), and the value of the argument (how far to move or turn). Their task was to predict the final orientation of the turtle, given its initial orientation and command.

Fay and Mayer first constructed an ideal model, comprised of about a dozen elementary operations. Then, based on the general characteristics of children's errors, they proposed six types of misconceptions (e.g., that a right-turn command actually slides the turtle to the right) and formulated models for the microstructure of each misconception, in terms of degenerate versions of relevant parts of the ideal model. For the subjects to which these degenerate models were applied, Fay and Mayer were able to account for nearly every one of the (mostly) incorrect responses to the 24 items in their test battery.

Error-analyses of this type are not only useful for cognitive developmental theory, but they also have pedagogical implications. The potential for facilitating remedial instruction is what originally motivated the Brown and Burton work on children's subtraction bugs, and it continues to be a valuable by-product of detailed error-analysis research:

> . . . novice Logo programmers appear to enter the Logo environment with individual confusions and misconceptions that they apply fairly consistently during instruction. Diagnosis of the specific confusions—such as a misunderstanding of what left and right mean or a misunderstanding of what degrees of rotation means—provides a more detailed and potentially useful evaluation of students' knowledge than the traditional global measurement of percentage correct. A cognitive diagnosis . . . provides information concerning *what* a student knows rather than a traditional measurement of *how much* a student can do. (Fay & Mayer, 1987, p. 265)

I believe that this kind of work illustrates the basic premise of this aspect of information-processing approaches: Careful and creative analysis of complex error patterns can provide an extremely informative window into the child's mental processes.

Analysis of Protocols

Protocol analysis is another form of high-density data that is often associated with information-processing approaches. The basic idea here is that, in addition to final responses on tasks, the subject can generate external indications of intermediate states, and that this pattern of intermediate indicators (the protocol) can be highly informative about the underlying processes that generated the final response. Included here are not only children's verbal protocols, such as the Siegler and Jenkins data described earlier, but also sequences of eye movements (Haith, 1980; Vurpillot, 1968) and other motor responses, such as reaching (Granrud, Haake, & Yonas, 1985). The classic verbal protocol analyses with adults are reported in Newell and Simon (1972), and a theoretical and methodological discussion of protocol analysis is offered in Ericsson and Simon (1984).

A common misconception about the verbal protocol analysis methodology is that it requires subjects to give an introspective account of their own behavior, and therefore is unreliable and unacceptably subjective (Nisbett & Wilson, 1977). Clearly, this would be a fatal flaw in the methodology, especially if it is to be used with children. But the criticism is unfounded. As Anderson (1987) summarized the issue:

> Many of these unjustified criticisms of protocols stem from the belief that they are taken as sources of psychological theory rather than as sources of data about states of the mind. For the latter, one need not require that the subject accurately interpret his mental states, but only that the theorist be able to specify some mapping between his reports and states of the theory. (p. 472)

In adult information-processing psychology, protocol analysis is a widespread method, but it is only infrequently used in more than a casual fashion by current cognitive developmentalists. This is very surprising, when one considers the fact that Piaget was the most prolific collector and analyzer of verbal protocols in the history of psychology.

Klahr and Robinson (1981) used a combination of motor and verbal protocol analysis and error analysis to explore preschool children's problem-solving and planning skills. They used a variant of the Tower of Hanoi puzzle (described earlier) in which both the initial state and the goal state were physically displayed (see Fig. 5.5). Children were presented with partially completed three-disk (actually, "three-can") problems requiring from two to seven moves to solution, and they were instructed to describe

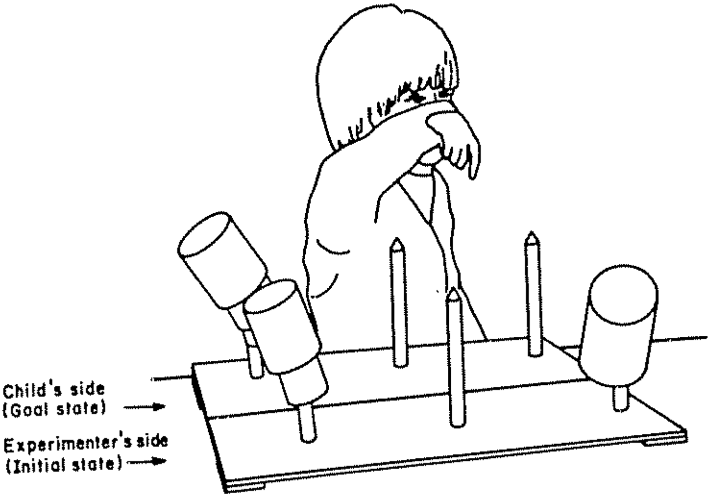


FIG. 5.5. Children's version of the Tower of Hanoi puzzle. The cans and pegs on the child's side of the table represent the goal state, and the can's and pegs on the opposite side represent the initial state. The child's task is to describe the full sequence of legal moves that will transform the initial state into the goal state. The problem depicted here can be solved in one move (from Klahr & Robinson, 1981).

the full sequence of moves that would change the initial state so that it matched the goal state. Children were videotaped as they described—verbally and by pointing—what sequence of moves they would use to solve the problem, but the cans were never actually moved. The protocols enabled Klahr and Robinson to infer the children's internal representation of the location of each can, and the processes whereby children made moves. They then constructed several alternative models of children's strategies, and used the error-analysis technique described earlier to identify each child's response pattern with a specific strategy. Note that nowhere were the children asked to reflect on their own mental processes, or to give a report on what strategies they were using while solving the problems.

The information extracted from the protocols in the Klahr and Robinson study consisted of a planned sequence of well-defined moves of discrete objects. This level of mapping from the protocol to hypothesized representations and processes is characteristic of the kind of protocol analyses presented in Newell and Simon's (1972) seminal work. A "richer" use of protocols, similar to some of the later examples in Ericsson and Simon (1984), provides the basis of recent investigations of children's strategies for scientific reasoning (Dunbar & Klahr; 1989; Kuhn, 1989; Schauble, 1990). Klahr, Fay, and Dunbar (1991) used verbal protocols as the primary data source in their study of experimentation strategies. Children (aged 8 to 11 years old) and adults were presented with a programmable robot, taught about most of its operating characteristics, and then asked to discover how some additional feature worked. They were asked to talk aloud as they generated hypotheses, ran experiments (i.e., wrote programs for the robot and ran them), and made predictions, observations, and evaluations. These verbal protocols were then analyzed in terms of different classes of hypotheses, the conditions under which experiments were run, how observed results were assessed, and so on. Based on this analysis, Klahr and his colleagues were able to characterize some of the differences in scientific reasoning skills between children and adults. In particular, they demonstrated that younger children have very poor general heuristics for designing experiments and evaluating their outcomes.

CONSTRAINTS AND LIMITATIONS

For all of its pervasiveness, the information-processing approach to cognitive development has several constraints and limitations. In her discussion of the limitations of information-processing approaches, Kuhn (this volume) addresses some of these issues. Here, I offer a somewhat different perspective. The extent to which these limitations and constraints are temporary or fundamental and permanent remains to be seen.

Populations

To date, hard-core information-processing approaches to cognitive development have been focused primarily on normal children over 2 years old. Those approaches that have dealt with younger, older or special populations have tended to be of the soft-core variety: for example, Davidson's (1986) study of gifted children, Hoyer and Familant's (1987) and Madden's (1987) studies of elderly subjects, and Geary, Widaman, Little, and Cormier's (1987) and Spitz and Borys' (1984) investigations of learning disabled and retarded subjects. In the case of special populations, issues are usually framed by the theoretical or empirical results emerging from studies of normal populations, and the question of interest is the qualitative or quantitative difference in a particular information-processing construct. For example, Spitz and Borys (1984) investigated the differences in search processes between normal and retarded adults on the classic Tower of Hanoi puzzle. Interesting work using information-processing concepts and techniques has also been done with non-human species. For example, Arbib (1987) proposed a series of models of visually guided behavior in frogs that utilize the schema notion described earlier. His work has many of the information-processing features discussed in this chapter, including formal representations and computer simulation. None of these populations, however, has been the subject of as much information-processing research as have typical children.

Topics

The developmental topics studied within this approach range from higher cognitive processes, such as problem solving (Resnick & Glaser, 1976) and scientific reasoning (Dunbar & Klahr, 1989; Kuhn & Phelps, 1982), to more basic processes, such as attention and memory (Chi, 1981; Kail, 1984). Because the focus of this chapter is cognitive development, I have drawn the conventional—and arbitrary—boundary that precludes an extensive discussion of perceptual-motor or language development. Nevertheless, I would be hard pressed to present a principled argument for excluding either of these areas from mainstream information processing, for in both of them one can find many examples of the approach (MacWhinney, 1987b; Yonas, 1988). MacWhinney's (1987a) edited volume on mechanisms of language acquisition contains an array of information-processing approaches that run the gamut from soft- to hard-core features. In the area of perceptual development, Marr's (1982) seminal work, which advocates computational models as the proper approach to constructing theories of vision, is increasingly influential. Indeed, Banks (1988), in presenting his own computational model of contrast constancy, argued that perceptual development is a more promising area in which to construct computational models than cognitive or social development, because there are more constraints that can be brought to bear to limit the proliferation of untested (and untestable) assumptions.

Change and Stability

Throughout this chapter I have offered examples of how information-processing approaches can account for cognitive changes. But stability is also an important feature of development and the topic has received extensive attention from developmentalists (Bornstein & Krasnegor, 1989). Developmentalists' interest in stability is primarily based on psychometric approaches. That is, it deals with relative constancy with respect to measures that compare one person to another. These approaches do not study the absence of change, but rather the lack of change in the rank order of individuals in a group on some measure. For example, consider the question of whether IQ scores are stable over time. IQ is a relative measure. The test is designed to order individuals from high to low, and any stability that is found in such scores means that if Sue scored higher than Sam at age 2, she continued to do so at age 4 (or 10 or 20). There is no question that both Sue and Sam have undergone substantial change in knowledge, skills, LTM structures, basic processes, and so forth. (Indeed, Kail's work, cited earlier, suggests that even at the level of elementary information processes, both Sue and Sam must have improved proportionally.)

The interesting question is why all of these changes in various aspects of the information-processing system have not changed the relative standing of Sue and Sam. To the best of my knowledge, the information-processing approaches that have addressed this question have either (a) been of the soft-core variety, or (b) focused on stability in infancy and early childhood (Bornstein, 1989; Colombo, Mitchell, O'Brien, & Horowitz, 1987; Fagan & Singer, 1983). A promising area for future research would be the application of microgenetic and chronometric techniques to individual differences and questions of change and stability.

Nonsymbolic Computational Architectures

One of the justifications given earlier for excluding perceptual, motor, and language development from this chapter was its focus on higher cognitive processes in children of school age and above. Another reason is that this chapter has focused on symbolically oriented information-processing approaches, to the exclusion of the newer connectionist framework. Advocates of this approach to computational models of cognition argue that information-processing approaches of the symbolic variety are inherently inadequate to account for the important phenomena in language acquisition and perceptual-motor behavior. The gist of the argument is that, given the highly parallel and "presymbolic" nature of these areas, it is doubtful that highly serial symbol-oriented information-processing models will ever be able to provide plausible accounts of development in these areas.

Indeed, this purported weakness is, according to some connectionists (McClelland, 1989; Rumelhart & McClelland, 1986), the Achilles heel of the symbolic approach to computational modeling. Furthermore, from a developmental perspective, the situation is particularly troublesome, for if we are to model a system from its neonatal origins, then we will have to invent new ways to model the interface between perceptual–motor systems and central cognition, particularly at the outset, when they provide the basis for all subsequent cognition.

Connectionism's advocates have suggested several very important and exciting possibilities, including the possibility that connectionist approaches may be particularly well suited for modeling biological changes underlying cognitive development. To date, the most interesting work has been in the area of language acquisition (MacWhinney, Leinbach, Taraban, & McDonald, 1989; Plunkett & Marchman, 1991), although there are a few connectionist models of higher order cognitive transitions, such as McClelland and

Jenkins' (1991) simulation of rule acquisition on Siegler's balance scale task. Many other connectionist models are summarized by Bechtel and Abrahamsen (1991). Included in their set of potential contributions are (a) a new interpretation of the distinction between maturation and learning, (b) a computational instantiation of the distinction between accommodation and assimilation, (c) an account of context effects (in which minor task variations have large effects on preschoolers' performance [Gelman, 1978]), and (d) explanations of many of the phenomena and anomalies associated with stages and transitions.

At present, there are not enough connectionist models of developmental phenomena to decide the extent to which they will replace, augment, or be absorbed by the symbolic variety of information-processing models described in this chapter. Nevertheless, both the broad-gauged connectionist criticisms of symbol-oriented approaches to cognition and the potential connectionist contributions to computational models of cognitive development warrant careful consideration.

CONCLUSIONS

Rather than attempt to summarize a chapter that is already a summary of ongoing research, in this concluding section I (a) reiterate the case for computational models of developmental phenomena, and (b) speculate about the future of information-processing approaches to cognitive development.

Why Bother?

Why should someone interested in theories of cognitive development be concerned about computational models of the sort discussed earlier? The primary justification for focusing on such systems is the claim that self-modification is the central question for cognitive developmental theory. It appears to me that in order to make theoretical advances, we will have to formulate computational models at least as complex as the systems described here.

Kuhn (Chap. 4, this volume) criticizes the information-processing approach for being insufficiently attentive to the issue of self-modification. As noted earlier, she is not alone in this regard, but there is some irony in the current situation. Although it is not difficult to find developmentalists who fault hard-core treatments of transition and change, it is even easier to find criticisms of the entire field of developmental psychology for its inability to deal adequately with these central topics.

> I have asked some of my developmental friends where the issue stands on transitional mechanisms. Mostly, they say that developmental psychologists don't have good answers. Moreover, they haven't had the answer for so long now that they don't very often ask the question anymore—not daily, in terms of their research. (Newell, 1990, p. 462)

Is this too harsh a judgment? Perhaps we can dismiss it as based on hearsay, for Newell himself is not a developmental psychologist. But it is harder to dismiss the following assessment from John Flavell (1984):

> . . . serious theorizing about basic mechanisms of cognitive growth has actually never been a popular pastime. . . . It is rare indeed to encounter a substantive treatment of the problem in the annual flood of articles, chapters, and books on cognitive development. The reason is not hard to find: Good theorizing about mechanisms is very, very hard to do. (p. 189)

Even more critical is the following observation on the state of theory in perceptual development from one of the area's major contributors in recent years:

> Put simply, our models of developmental mechanisms are disappointingly vague. This observation is rather embarrassing because the aspect of perceptual developmental psychology that should set it apart from the rest of perceptual psychology is the explanation of how development occurs, and such an explanation is precisely what is lacking. (Banks, 1987, p. 342)

It is difficult to deny either Newell's or Bank's assertions that we don't have good answers, or Flavell's assessment of the difficulty of the question. However, I believe that it is no longer being avoided: Many developmentalists have been at least asking the right questions recently. In the past few years we have seen Sternberg's (1984) edited volume *Mechanisms of Cognitive Development,* MacWhinney's (1987b) edited volume *Mechanisms of Language Acquisition,* and Siegler's (1989) *Annual Review* chapter devoted to transition mechanisms. So the question is being asked.

Furthermore, the trend is in the direction of hardening the core. Only a few of the chapters in the 1984 Sternberg volume specify mechanisms any more precisely than at the flow-chart level, and most of the proposed mechanisms are at the soft end of the information-processing spectrum. However, only 3 years later, Klahr et al.'s (1987) *Production System Models of Learning and Development* included several chapters that described running programs, and within 5 years, Siegler (1989), in characterizing several general categories for transition mechanisms (neural mechanisms, associative competition, encoding, analogy, and strategy choice), cited computationally based exemplars for all but the neural mechanisms (e.g., Bakker & Halford, 1988; Falkenhainer, Forbus, & Gentner, 1986; Holland, 1986; MacWhinney, 1987a; Rumelhart & McClelland, 1986; Siegler, 1988).

A clear advantage of such computational models is that they force difficult questions into the foreground, where they can be neither sidetracked by the wealth of experimental results, nor obscured by vague characterizations of the various essences of cognitive development. The relative lack of progress in theory development—noted by Banks, Flavell, and Newell—is a consequence of the fact that, until recently, most developmental psychologists have avoided moving to computationally based theories, attempting instead to attack the profoundly difficult question of self-modification with inadequate tools. Mastery of the new tools for computational modeling is not easy. Nevertheless it appears to be a necessary condition for advancing our understanding of cognitive development. As Flavell and Wohlwill (1969) noted more than 20 years ago: "Simple models will just not do for human cognition" (p. 74).

The Future of the Hard-Core Approach

That brings me to my second concluding topic: the education of future cognitive developmentalists. The conceptual and technical skills necessary for computational modeling require training of a different sort than one finds in most graduate programs today. However, I see the current situation as analogous to earlier challenges to the technical content of graduate training. When other kinds of computational technology that are now in common use—such as statistical packages, or scaling procedures—were first being applied to psychological topics, journal articles invariably included several pages of description about the technique itself. Writers of those early articles correctly assumed that their readers needed such background information before the psychological issue of interest could be addressed. Today, writers of papers using analysis of variance, or multidimensional scaling, or path analysis simply assume that their readers have had several courses in graduate school, learning the fundamentals.

Similarly, in the early years of computer simulation, the necessary resources of large main frame computers were limited to very few research centers, and exposure to computational modeling was inaccessible to most developmentalists. Even today, few developmental psychologists have had any training with computational models, and only a handful of computational modelers have a primary interest in cognitive development. Nevertheless, the intersection of these two areas of research is growing. (The 1991 meeting of the Society for Research in Child Development included two hard-core symposia, one entitled "Connectionist Models and Child Development" and the other "Computational Models of Cognitive Transition

Mechanisms.") Moreover, with the increasing availability of powerful work-stations, the proliferation of computer networks for dissemination of computational models, and the increasing number of published reports on various kinds of computationally based cognitive architectures, the appropriate technology and support structures are becoming widely accessible. This accessibility will make it possible to include simulation methodology as a standard part of graduate training.

My hope is that, over the next few decades, we will begin to see many papers about cognitive development couched in terms of extensions to systems like Soar, or ACT*, or some other well-known (by then) cognitive architecture, or some future connectionist model. Just as current writers need not explain the conceptual foundations of an analysis of variance, so future writers will deem it unnecessary to include tutorials on computational models in their papers. Once we are fully armed with such powerful tools, progress on our most difficult problems will be inevitable. We will no longer talk of approaches to our problems, but rather, of proposals for their solutions.

ACKNOWLEDGMENTS

This chapter is an adaptation and extension of an earlier characterization of hard-core and soft-core approaches (Klahr, 1989) and includes some issues originally treated in Neches, Langley, and Klahr (1987). Preparation of the chapter was supported in part by a grant from the A. W. Mellon Foundation. Very helpful comments and suggestions for improvements in an earlier draft came from Robert Siegler, Leona Schauble, and David Penner. Becky Duxbury provided excellent secretarial support throughout the project.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192–210.

Anderson, J. R., Kline, P. J., & Beasley, C. M., Jr. (1978). *A general learning theory and its application to schema abstraction* (Tech. Rep. 78-2). Pittsburgh, PA: Carnegie Mellon University, Department of Psychology.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86,* 124–140.

Arbib, M. A. (1987). Levels of modeling of mechanisms of visually guided behavior. *Behavioral and Brain Sciences, 10,* 407–465.

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in* *developmental psychology: Progress in cognitive development research* (pp. 302–338). New York: Springer-Verlag.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 90–195). New York: Academic Press.

Bakker, P. E., & Halford, G. S. (1988). *A basic computational theory of structure-mapping in analogy and transitive inference* (Tech. Rep.). St. Lucia, Australia: University of Queensland, Centre for Human Information Processing and Problem Solving.

Banks, M. S. (1987). Mechanisms of visual development: An example of computational models. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 339–371). New York: Springer-Verlag.

Banks, M. S. (1988). Visual recalibration and the development of contrast and optical flow perception. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 145–196). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barrett, T. R. (1978, November). *Aging and memory: Declining or differences.* Paper presented at the meeting of the Psychonomic Society, San Antonio, TX.

Baylor, G. W., & Gascon, J. (1974). An information processing theory of aspects of the development of weight seriation in children. *Cognitive Psychology, 6,* 1–40.

Baylor, G. W., Gascon, J., Lemoyne, G., & Pother, N. (1973). An information processing model of some seriation tasks. *Canadian Psychologist, 14,* 167–196.

Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the mind.* Cambridge, MA: Basil Blackwell.

Beilin, H. (1983). The new functionalism and Piaget's program. In E. K. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory?* (pp. 3–40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bornstein, M. H. (1989). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M. H. Bornstein & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives* (pp. 147–170). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bornstein, M. H., & Krasnegor, N. A. (Eds.) (1989). *Stability and continuity in mental development: Behavioral and biological perspectives.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brainerd, C. J. (1987). Structural measurement theory and cognitive development. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 1–37). New York: Springer-Verlag.

Brown, A. L. (1982). Learning and development: The problem of compatibility, access and induction. *Human Development, 25,* 89–115.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2,* 155–192.

Brown, J. S., & VanLehn, K. (1982). Towards a generative theory of 'bugs'. In T. Romberg, T. Carpenter, & J. Moser (Eds.), *Addition and subtraction: A developmental perspective* (pp. 117–135). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, R. (1970). Introduction. In Society for Research in Child Development (Ed.), *Cognitive development in children* (pp. ix–xii). Chicago: University of Chicago Press.

Burt, C. (1919). The development of reasoning in school children. *Journal of Experimental Pedagogy, 5,* 68–77.

Card, S., Moran, T. P., & Newell, A. (1983) *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carver, S. M. (1986). *Transfer of LOGO debugging skill: Analysis, instruction and assessment.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Case, R. (1985). *Intellectual development: Birth to adulthood.* New York: Academic Press.

Case, R. (1986). The new stage theories in intellectual development: Why we need them; what they assert. In M. Perlmutter (Ed.), *Perspectives for intellectual development* (pp. 57–91). Hillsdale, NJ: Lawrence Erlbaum Associates.

Case, R., & Griffin, S. (1990). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. In C. A. Hauert (Ed.), *Developmental psychology: Cognitive, perceptuo-motor and neuropsychological perspectives* (pp. 193–230). Amsterdam: North Holland.

Chase, W. G. (Ed.). (1973). *Visual information processing*. New York: Academic Press.

Chase, W. G. (1978). Elementary information processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 19–90). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition, 4,* 559–572.

Chi, M. T. H. (1977). Age differences in memory span. *Journal of Experimental Child Psychology, 23,* 266–281.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (1981). Knowledge development and memory performance. In M. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 221–229). New York: Plenum.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. *Advances in Child Development and Behavior, 20,* 91–141.

Chi, M. T. H., & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology, 19,* 434–439.

Colombo, J., Mitchell, D. W., O'Brien, M., & Horowitz, F. D. (1987). The stability of visual habituation during the first year of life. *Child Development, 58,* 474–487.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Davidson, J. E. (1986). The role of insight in giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 201–222). New York: Cambridge University Press.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science, 238,* 1556–1557.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery strategies. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109–143). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ericsson, A. & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fagan, J. F., & Singer, L. T. (1983). Infant recognition memory as a measure of intelligence. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 2, pp. 31–78). Norwood, NJ: Ablex.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1986). The structure-mapping engine. In 5th *Proceedings of the American Association for Artificial Intelligence* (pp. 272–277). Philadelphia: American Association for Artificial Intelligence.

Fay, A. L., & Mayer, R. E. (1987). Children's naive conceptions and confusions about LOGO graphics commands. *Journal of Educational Psychology, 79*(3), 254–268.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87,* 477–531.

Flavell, J. H. (1984). Discussion. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 187–210). New York: Freeman.

Flavell, J. H., & Wohlwill, J. F. (1969). Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development* (pp. 67–120). New York: Oxford University Press.

Geary, D. C., Widaman, K. F., Little, T. D., & Cormier, P. (1987). Cognitive addition: Comparison of learning disabled and academically normal elementary school children. *Cognitive Development, 2*(3), 249–270.

Gelman, R. (1978). Cognitive development. *Annual Review of Psychology, 29,* 297–332.

Gelman, R. (1982). Accessing one-to-one correspondence: Still another paper about conservation. *British Journal of Psychology, 73,* 209–220.

Goodman, N. (1968). *Languages of art.* Indianapolis: Bobbs-Merrill.

Granrud, C. E., Haake, R. J., & Yonas, A. (1985). Infants' sensitivity to familiar size: The effect of memory on spatial perception. *Perception and Psychophysics, 37,* 459–466.

Gagné, R. M. (1968). Learning hierarchies. *Educational Psychology, 6,* 1–9.

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16*(1), 94–143.

Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology, 4,* 246–276.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79,* 329–343.

Haith, M. M. (1980). *Rules that infants look by.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Halford, G. S. (1970). A theory of the acquisition of conservation. *Psychological Review, 77,* 302–316.

Halford, G. S. (1975). Children's ability to interpret transformations of a quantity, I: An operational system for judging combinations of transformations. *Canadian Journal of Psychology, 29,* 124–141.

Hill, J. C. (1983). A computational model of language acquisition in the two-year-old. *Cognition and Brain Theory, 6,* 287–317.

Hill, J. C., & Arbib, M. A. (1984). Schemas, computation, and language acquisition. *Human Development, 27,* 282–296.

Holland, J. H. (1986). Escaping brittleness: The possibilities of general purpose machine learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 593–624). Los Altos, CA: Morgan-Kaufmann.

Hoyer, W. J., & Familant, M. E. (1987). Adult age differences in the rate of processing expectancy information. *Cognitive Development, 2*(1), 59–70.

Hunter, I. M. L. (1968). Mental calculation. In P. C. Wason & P. N. Johnson-Laird (Eds.), *Thinking and reasoning* (pp. 341–351). Baltimore, MD: Penguin.

Inhelder, B., Ackerman-Vallado, E., Blanchet, A., Karmiloff-Smith, A., Kilcher-Hagedorn, H., Montagero, J., & Robert, M. (1976). The process of invention in cognitive development: A report of research in progress. *Archives de Psychologie, 171,* 57–72.

Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology, 30,* 363–392.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3,* 281–282.

Kail, R. (1984). *The development of memory in children* (2nd ed.). New York: Freeman.

Kail, R. (1986). Sources of age differences in speed of processing. *Child Development, 57,* 969–987.

Kail, R. (1988). Developmental functions for speeds of cognitive processes. *Journal of Experimental Child Psychology, 45,* 339–364.

Kail, R. (1991a). Processing time declines exponentially during childhood and adolescence. *Developmental Psychology, 27,* 259–266.

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

Karmiloff-Smith, A. (1984). Children's problem solving. In M. Lamb, A. Brown, & B. Rogoff (Eds.), *Advances in developmental psychology* (pp. 33–99). Hillsdale, NJ: Lawrence Erlbaum Associates.

Keating, D. P., & Bobbitt, B. L. (1978). Individual and developmental differences in cognitive processing components of mental ability. *Child Development, 49,* 155–167.

Kellman, P. H. (1988). Theories of perception and research in perceptual development. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 267–282). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem solving. *Psychological Review, 69,* 1–16.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109–129.

Klahr, D. (1973). A production system for counting, subitizing, and adding. In W. G. Chase (Ed.), *Visual information processing* (pp. 3–34). New York: Academic Press.

Klahr, D. (1989). Information processing approaches. In R. Vasta (Ed.), *Annals of child development* (pp. 131–185) Greenwich, CT: JAI Press.

Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20,* 362–404.

Klahr, D., Fay, A. L., & Dunbar, K. (1991). *Heuristics for scientific experimentation: A developmental study.* Working Paper, Carnegie Mellon University.

Klahr, D., Langley, P., & Neches, R. (Eds.). (1987). *Production systems models of learning and development.* Cambridge, MA: MIT Press.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem solving and planning processes in preschool children. *Cognitive Psychology, 13,* 113–148.

Klahr, D., & Siegler, R. S. (1978). The representation of children's knowledge. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 12, pp. 61–116). New York: Academic Press.

Klahr, D., & Wallace, J. G. (1970a). The development of serial completion strategies: An information processing analysis. *British Journal of Psychology, 61,* 243–257.

Klahr, D., & Wallace, J. G. (1970b). An information processing analysis of some Piagetian experimental tasks. *Cognitive Psychology, 1,* 358–387.

Klahr, D., & Wallace, J. G. (1972). Class inclusion processes. In S. Farnham-Diggory (Ed.), *Information processing in children* (pp. 144–181). New York: Academic Press.

Klahr, D., & Wallace, J. G. (1973). The role of quantification operators in the development of conservation of quantity. *Cognitive Psychology, 4,* 301–327.

Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information processing view.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kolers, P. A., & Smythe, W. E. (1984). Symbol manipulation: Alternatives to the computational view of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 289–314.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific reasoning skills.* Orlando, FL: Academic Press.

Kuhn, D., & Phelps, E. (1982). The development of problem solving strategies. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 2–44). New York: Academic Press.

Lachman, R., Lachman, J., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J. H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 311–334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lewis, C. H. (1978). *Production system models of practice effects.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lindberg, M. A. (1980). Is knowledge base development a necessary and sufficient condition for memory development? *Journal of Experimental Child Psychology, 30,* 401–410.

Mackworth, A. K. (1987). What is the schema for a schema? *Behavioral and Brain Sciences, 10*(3), 443–444.

MacWhinney, B. J. (1987a). Competition and cooperation in language processing. In R. Tomlin (Ed.), *Proceedings of the Pacific Conference on Linguistics.* Eugene, OR: University of Oregon Press.

MacWhinney, B. J. (Ed.). (1987b). *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28,* 255–277.

Madden, D. J. (1987). Aging, attention, and the use of meaning during visual search. *Cognitive Development, 2,* 201–216.

Mandler, J. M. (1983). Representation. In P. H. Mussen (Ed.), *Cognitive development: Vol. III. Handbook of child psychology* (pp. 420–494). New York: Wiley.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111–151.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 8–45). Oxford: Clarendon Press.

McClelland, J. L., & Jenkins, E. (1991). Nature, nurture, and connections: Implications of connectionist models for cognitive development. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 41–73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Metz, K. E. (1985). The development of children's problem solving in a gears task: A problem space perspective. *Cognitive Science, 9,* 431–471.

Miller, P. H. (1983). *Theories of developmental psychology.* San Francisco: Freeman.

Neches, R. (1981). *Models of heuristic procedure modification.* Unpublished doctoral dissertation, Department of Psychology, Carnegie Mellon University.

Neches, R. (1987). Learning through incremental refinement of procedures. In D. Klahr, P. Langley & R. Neches (Eds.), *Production system models of learning and development* (pp. 163-220). Cambridge, MA: MIT Press.

Neches, R., Langley, P., & Klahr, D. (1987). Learning, development and production systems. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 1–53). Cambridge, MA: MIT Press.

Nelson, K., & Gruendel, J. M. (1981). Generalized event representations: Basic building blocks of cognitive development. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. 1, pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum Associates.

Neves, D., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A. (1966). On the representation of problems. *Computer science research review,* 18–33. Computer Science Department, Carnegie Institute of Technology.

Newell, A. (1972). A note on process-structure distinctions in developmental psychology. In S. Farnham-Diggory (Ed.), *Information processing in children* (pp. 126–143). New York: Academic Press.

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 463–526). New York: Academic Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4,* 135–183.

Newell, A. (1981). Reasoning problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 693–718). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. (1975). *Explorations in cognition*. San Francisco: Freeman.

Palmer, S. E., & Kimchi, R. (1986). The information processing approach to cognition. In T. J. Knapp & L. C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies* (pp. 37–77). Hillsdale, NJ: Lawrence Erlbaum Associates.

Piaget, J. (1928). *Judgement and reasoning in the child*. New York: Harcourt, Brace.

Piaget, J. (1929). *The child's conception of the world*. New York: Harcourt, Brace.

Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a back-propagation network: Implications for language acquisition. *Cognition, 38,* 1–60.

Rabinowitz, F. M., Grant, M. J., & Dingley, H. L. (1987). Computer simulation, cognition, and development: An introduction. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 263–301). New York: Springer-Verlag.

Resnick, L. B., & Glaser, R. (1976). Problem solving and intelligence. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 205–230). Hillsdale, NJ: Lawrence Erlbaum Associates.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–200). New York: Academic Press.

Rosenbloom, P. S., & Newell, A. (1982). Learning by chunking: Summary of a task and a model. In *Proceedings of the Second National Conference on Artificial Intelligence* (pp. 255–257). Los Altos, CA: Morgan-Kaufmann.

Rosenbloom, P. S., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 221–286). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31–57.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*(4), 481–520.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46* (Whole No. 189).

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250–264.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill *Journal of Experimental Psychology: General, 117,* 258–275.

Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology, 117,* 258–275.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46,* 606–620.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325–348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H. A. (1962). An information processing theory of intellectual development. *Monographs of the Society for Research in Child Development, 27,* (2, Serial No. 82).

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Simon, T., Newell, A., & Klahr, D. (1991). A computational account of children's learning about number conservation. In D. Fisher & M. Pazzani (Eds.), *Concept formation: Knowledge and experience in unsupervised learning* (pp. 423–462). Palo Alto, CA: Morgan-Kauffman.

Spitz, H. H., & Borys, S. V. (1984). Depth of search: How far can the retarded search through an internally represented problem space? In P. H. Brooks, R. Sperber, & C. McCauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 333–358). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing,* (Vol. 2., pp. 53–120). Norwood, NJ: Ablex.

Sternberg, R. J. (Ed.). (1984). *Mechanisms of cognitive development.* New York: Freeman.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology, 27,* 195–232.

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

Strauss, S. (Ed.). (1982). *U-shaped behavioral growth.* New York: Academic Press.

Townsend, J. T., & Kadlec, H. (1990). Psychology and mathematics. In R. E. Mickens (Ed.)., *Mathematics and science* (pp. 223–249). Singapore: World Scientific Publishing.

Trabasso, T. (1975). Representation, memory and reasoning: How do we make transitive inferences? In A. D. Pick (Ed.), *Minnesota symposia on child psychology.* (Vol. 9., pp. 135–172). Minneapolis: University of Minnesota Press.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. J. Falmagne (Ed.), *Reasoning: Representation and process* (pp. 201–230). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology, 6,* 632–650.

Wallace, J. G., Klahr, D., & Bluff, K. (1987). A self-modifying production system of cognitive development. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 356–435). Cambridge, MA: MIT Press.

Wertsch, J. V., & Hickmann, M. (1987). Problem solving in social interaction: A microgenetic analysis. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 251–266). New York: Academic Press.

Wilkinson, A. C., & Haines, B. A. (1987). Learning a cognitive skill and its components. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 86–114). New York: Springer-Verlag.

Yonas, A. (Ed.). (1988). *Perceptual development in infancy.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Young, R. M. (1976). *Seriation by children: An artificial intelligence analysis of a Piagetian task.* Basel: Birkhauser.

Young, R. M., & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science, 5,* 153–177.